TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

TEEMU ESKELINEN
ORDER FLOW CHARACTERISTICS IN NORDIC STOCK EXCHANGES

Master of Science thesis

Examiner: Prof. Juho Kanniainen
Examiner and topic approved by the
Faculty Council of the Faculty of
Business and Built Environment
on 8th April 2015

ABSTRACT

TEEMU ESKELINEN: Order Flow Characteristics in Nordic Stock Exchanges
Tampere University of Technology
Master of Science thesis, 55 pages
August 2015
Master's Degree Programme in Industrial Management
Major: Industrial Engineering
Examiner: Prof. Juho Kanniainen
Keywords: order flow, market microstructure, stochastic order book modeling

This thesis studied the statistical characteristics of order flow in Nordic Stock Exchanges. These were also reflected against the recent academic literature in theoretical order flow dynamics. The theoretical and empirical aspects of limit order book markets were explored in a short literature review.

The main focus of the study were the equity exchanges in Helsinki, Stockholm and Copenhagen aspects of which were explored in the empirical part of the thesis. The data set used was NASDAQ OMX Nordic TotalView ITCH feed. Order sizes, inter-arrival times, price impacts and relative positions were extracted from this data set for the period June 2010 – May 2013. These factors are important in, for example, calibrating mathematical order book models for the Nordic Stock Exchanges.

The order flow characteristics mainly followed the outlines set in previous studies conducted in exchanges around the world. In Nordic Exchanges, the order flow was divided into a few very high liquidity stocks and rest of lower liquidity, which made the reporting of generalized descriptive constants difficult. The diurnal seasonality was also found to be very significant and was affected by e.g. the daily openings of U.S. exchanges between 14:00 – 15:00 CET. The relative price power law decay was found to hold with exponents $\alpha_{hel} = 1.15$, $\alpha_{sto} = 1.18$ and $\alpha_{cph} = 1.20$.

Additionally to these descriptive statistics, the study tested the assumptions made by the recent mathematical models. The assumptions were discovered to be too simplistic and e.g. the usage of homogeneous Poisson processes to govern the arrival of order book events was problematic in long time spans. Compared to the Cont et al. (2010) arrivals were reported to be clustered in time, with significant short-term autocorrelation. The arrival times showed humps around 20 milliseconds. The usage of Weibull distributions as total arrival rate process and the Order Flow Imbalance as a short term estimator of immediate price impact were shown to be appropriate for high liquidity stocks with $R^2 > 0.40$.

TIIVISTELMÄ

TEEMU ESKELINEN: Pohjoismaisten pörssien tilausvirtojen ominaisuudet
Tampereen teknillinen yliopisto
Diplomityö, 55 sivua
Elokuu 2015
Tuotantotalouden koulutusohjelma
Pääaine: teollisuustalous
Tarkastajat: Prof. Juho Kanniainen
Avainsanat: tilausvirta, markkinoiden mikrostruktuuri, stokastiset tilauskirjamallit

Tässä diplomityössä tutkittiin pohjoismaisten pörssien tilausvirtojen tilastollisia ominaisuuksia teoreettisien tilauskirjadynamiikkojen valossa. Työ käy läpi myös lyhyen kirjallisuuskatsauksen rajahinnallisten tilauskirjojen mallinnukseen ja ominaisuuksiin. Tutkimuksen kohteena olevat arvopaperipörssit sijaitsivat Helsingissä, Tukholmassa ja Kööpenhaminassa. Empiirinen osuus tarkasteli näiden ominaisuuksia käytten TotalView ITCH tietokantaa kolmen vuoden ajalta 2010 – 2013. Tilausvirtadatasta selvitettiin muun muassa tilausten koot, niiden välinen aika, hintavaikutus sekä suhteellinen paikka tilauskirjoissa.

Tilauskirjojen tilastolliset ominaisuukset seurasivat pääosin aiemmin kirjallisuudessa havaittuja lainalaisuuksia. Pohjoismaisten pörssien keskittyminen muutamaan likvideetiltään korkeaan osakkeeseen hankaloitti tulosten yleistävyyttä. Tilausvirtojen kausivaihtelu päivän sisällä oli myös erittäin voimakasta ja siihen vaikutti myös esimerkiksi USA:n pörssien avautuminen päivittäin aikavälillä 14:00 – 15:00 CET. Esimerkiksi potenssilain todettiin sopivan suhteellisten hintojen kumulatiiviseen jakaumaan pörssikohtaisin eksponentein $\alpha_{hel} = 1.15$, $\alpha_{sto} = 1.18$ ja $\alpha_{cph} = 1.20$.

Työssä selvitettiin myös matemaattisten mallien sopivuutta ja niiden oletusten paikkaansapitävyyttä suhteessa oikeisiin tilauskirjamarkkinoihin. Lopputuloksena nousi esille homogeenisiin Poisson prosesseihin perustuvien mallien oletusten olevan liian yksinkertaisia verrattuna oikeisiin tilauskirjoihin. Saapumisten osoitettiin olevan ajallisesti klusteroituneita ja niillä olevan merkittävä lyhyen aikavälin autokorrelaatio. Erityisen ongelmallista malleissa oli tilauskirjatapahtumien saapumisnopeuksien mallintaminen vakiosuuruisena. Toisaalta Weibulljakauman käyttö korkean vaihtoasteen osakkeiden kokonaissaapumisjakaumana ja tilausvirtaepätasapainon käytön lyhyen aikavälin hintavaikutusestimaattorina todettiin toimivaksi ($R^2 > 0.40$).

PREFACE

This thesis was written during the Spring of 2015 after I was introduced into the subject by Professor Juho Kanniainen. Additionally to his help, I'd like to acknowledge Jaakko Valli, who assisted me in handling the vast amount of data. Big thanks also goes to the authors of open source data analysis software, whose contributions made it possible to complete the thesis in such short time span. For her support and motivation during this hectic spring, I thank my girlfriend Tytti.

Tampere, 26.7.2015

Teemu Eskelinen

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

LOB	Limit Order Book
DMA	Direct Market Access
GTC	Good Till Close
GTT	Good Till Time
FOK	Fill-or-Kill
TAQ	Trades and Quotes dataset
OFI	Order Flow Imbalance
ZI	Zero Intelligence
DE	Dynamic Equilibrium
CST-model	Cont, Stoikov and Talreija (2010) model of LOB dynamics

δ	tick size
σ	lot size
Δ	relative price
P	(monetary) price
$\mathcal{L}$	order book state

1. INTRODUCTION

A well over half of current equity trading in the world is done on electronic limit order book markets (Jain 2003). In these venues, instead of trading with dealers in a Request For Quote (RFQ) markets, the participants can submit (limit or market) sell or buy orders which are matched with each other according to the exchange's market model. This combination of market orders, limit orders, trades and cancellations are *market events* and they are the building blocks of the *order flow*.

This change from market maker and dealer driven exchanges to a computerized systems has enabled the study at the level of a single trade or an order. Ability to submit, change and execute orders in a rapid pace gives rise to complex interplay with new orders and the state of the limit order book. This has been increasingly analyzed by both empirical and theoretical methods during the past 10 years. Many important observations in real life limit order books have been made that have required new models for order book dynamics to be developed.

This Master's thesis attempts to explore the main characteristics of order flow of single stocks and, by extension, the whole market in Nordic Stock Exchanges run by NASDAQ OMX Group. The combined NASDAQ OMX Nordic Exchange is the 16th biggest stock exchange in the world measured by the combined market capitalization and was created by OMX acquisition by NASDAQ in 2008. It consists of four distinct exchanges in Helsinki, Copenhagen, Stockholm, Reykjavik, Tallinn, Riga and Vilnius. These exchanges use the INET trading platform in which all order flow is electronic.

This study examines three of these exchanges: Stockholm, Copenhagen and Helsinki. In this thesis, the names NASDAQ OMX Nordic, OMX Nordic and the Nordic Stock exchanges are used interchangeably to reference to this set of three limit order markets. Many listed companies and market participants have common elements due to historical connections between the Nordic Countries. The equity trading between these three countries has become more and more interconnected and multiple stocks are cross listed in two or more participating exchanges. This gives rise to a interesting and comprehensive study subject for testing the stylized facts and theoretical models presented in the recent academic literature.

1.1 Background and research objectives

The objective of this thesis consists of two interconnected parts. Firstly, the recent academic literature on limit order book market modeling is explored and their most important calibration values (or characteristics) are itemized. Secondly, the paper attempts to describe these order and market level characteristics of the OMX Nordic limit order book market. These values include seasonality effects, order lifetime, inter-arrival times and price impact. Additionally the properties such as event clustering and Order Flow Imbalance as a measure of short term volatility are explored.

The theories tested and models implemented rely greatly on empirical and theoretical work done on limit order books over the past ten years. Most of the empirical studies conducted previously have focused on larger exchanges or used less detailed data sets of much shorter periods. This study's sample covers three years worth of high quality data at the technical implementation level, which enables exact examination of various order level phenomena. Due to the size of the data set, the thesis uses visualization and descriptive tables to illustrate the most important aspects of the study.

Apart from descriptive sections on limit order flow characteristics, the paper examines the differences between distinct exchanges inside NASDAQ OMX Nordic. The results are moreover reflected on previous empirical work on other exchanges by the use of literary review. The facts discussed in the review portion and explored in empirical analysis are important for creating robust limit order book models. Any model that fails to reproduce these *stylized facts* can not be considered to be satisfactory. Stylized facts have been observed in a large amount of limit order book markets trading multiple security classes (Bisiere and Kamionka 2000; Al-Suhaibani and Kryzanowski 2000; Lo et al. 2002; Zovko and Farmer 2002; Abhyankar et al. 1997). Recently there has been an attempt to collect and formally define the facts observed in the literature. Such listing has been provided by Gould et al. (2013). This thesis does not attempt completely complement and follow the listing made by Gould et al., but the paper has been very influential in the choice of the facts studied.

While some these statistical aspects of order flow have been studied in U.S. ITCH data before (Hollifield et al. 2004), this thesis explores and reports a number of unique and contemporary characteristics for the Nordic Stock Exchanges. Compared to past empirical work done on limit order books that have focused on very large exchanges in London, New York or Paris, this thesis leverages the current academic literature to comprehensively test the reported phenomena in smaller, open and fully electronic exchanges in "periphery" markets with very concentrated liquidity and lower number of participants.

Additionally, the thesis attempts to bridge the empirical and theoretical work done on the limit order book markets done in past ten years. The theoretical models can be divided into two distinct categories of Dynamic Equilibrium and Zero Intelligence models. Dynamic equilibrium describes the limit order book market heading for an equilibrium by an interplay of perfectly rational actors trying to maximize their personal utility functions. Most influential of these models have been proposed by Foucault et al. (2005) and Rosu (2014). Opposite of these are the Zero Intelligence approaches, where the order flow is modeled as a Markovian point process. Recent developments on the ZI models have been made by Cont et al. (2010), Zhao (2010) and Toke (2011). Besides of the dynamic equilibrium and zero intelligence models that attempt to describe the general dynamics of limit order book markets, there have been more limited, specialized models, which have focused on short time span applications. Most relevant to the study of *order flow* is the Order Flow Imbalance model of price impact introduced by Cont et al. (2014).

The comparative theoretical approach mainly focuses on the Stochastic Model for Limit Order Book Dynamics presented by Cont et al. (2010). The model was chosen due to its notability, influence and good fit with the data set used. The model has been built upon and improved by later authors, such as Zhao (2010), who have attempted to resolve the inconstancies that are explored in e.g. this thesis. The models have taken different approaches into solving the problems and no single model has been yet deemed significantly superior to another. This thesis does not attempt to "rank" these newer developments, but specify empirical factors that should be taken into account when formulating these models. As such, it lies in between the purely empirical and theoretical papers presenting an overview of their interrelations and future requisites. While the problems pointed out in current theoretical models have been made previously by other authors (Zhao 2010; Toke 2011) and somewhat acknowledged by the model authors themselves, provides this thesis more evidence using a larger and a more detailed data set from a different exchanges than the previous studies.

Basic theoretical background to limit order books and their modelling is provided in Chapter 2 while the Chapter 3 examines the literature on specific aspects in order flow and order execution that can be studied further using empirical methods. The Chapter 5 leverages these models to extract important and interesting values from NASDAQ OMX ITCH feed. These can be used to characterize order flow in general and calibrate theoretical limit order book models. Conclusions on special aspects of Nordic Stock Exchanges and their theoretical implications are discussed in Chapter 6.

1.2 Caveats and limitations

Due to the size of the data set, the number of securities chosen for further inspection was constrained. What is more, some securities did not have enough limit order activity to allow descriptive statistical analysis or modeling to take place. For these reasons, the generalisability of the results must be critically assessed before drawing e.g. exchange wide facts. On the other hand, the sheer time span and detail level of the sample can help attaining reasonable fits for statistical models, in spite of short term noise and outliers.

Other significant source of possible error is the nature of the data set. The TotalView ITCH feed provides a *technical* representation of the limit order book events. From this data source it's impossible derive the precise types of limit orders or their additional provisions. Some atomic actions by market agents such as changing the price of an order are indistinguishable from combinations of other order book events (in this case simultaneous order cancellation and insertion). To examine these factors, an access to proprietary broker level databases would be needed and such studies have been conducted before. Arising limitations are discussed further in Chapter 4.

In other words, the specific reasons agents post and cancel orders are not discussed. For the majority of the document, the order flow is considered as given, affected only by itself and the general state of the market. This assumption is later apparent in choice of models and facts explored.

2. ELECTRONIC LIMIT ORDER BOOK MARKET

This chapter provides the basic theoretical model of an stylized limit order book market without special order types or opening / closing auctions. The order book is presented as a interplay of queuing limit orders, which leave the queue either by *cancellations* or *market orders*. Moreover, a short review of empirical aspects and theoretical modeling of limit order book markets are outlined.

The stylized model represented here is later complemented with the more complicated OMX Nordic's real life model specified in the Chapter 4.

2.1 Order book

In a general model of limit order book market the orders are posted to a visible medium and specifying their price and quantity. In an electronic limit order books, this visible medium is usually an interconnected computer system. The system performs the order matching and distributes the state of the limit order books to all market participants with a *Direct Market Access* (DMA).

These participants can be e.g. brokerage firms or trading houses, who have access to the full limit order book state. These participants can further give access to the LOB to their clients, who can post orders through them, while usually having only partial information on the state of the order book. Even Direct Market Access participants are subject to *lag* or *latency* introduced by the electronic communications, measured usually in the range of tens of milliseconds. This means that in high traffic order books, the state has usually already been altered by an another event by the time the order book state update reaches the participants. This creates a bubble of inconsistency between market actors, which can be a significant factor in some trading strategies requiring very fast order placement or very precise knowledge of the whole order book. (Menkveld and Zoican 2014)

2.2 Limit orders

The main building block of LOB markets are *limit orders*, which provide the market with liquidity. Limit orders define both the *limit price*, which is the worst price that the order can be executed at, and the *quantity*, which is the amount of securities (shares) that the participant is willing to buy or sell.

More formally, limit orders are ex ante pre-commitments (t, j, x, p) to trade security j up to the quantity x at the price of p at time t. Limit orders are not, in theoretical sense, executed immediately (otherwise they would be classified as market orders) and the outstanding limit orders create together the current order book state, which explored further in the Section 2.4. In the academic literature and in this thesis, the buy and sell side limit orders are simply referred to as *bids* and *asks*. (Kukanov 2013)

The limit order book can be thus presented as a set of queues consisting of sell and buy orders with a corresponding price. These orders are called *active orders*. Before matching they lie in the incoming queue as *passive* orders. Since the matching is essentially immediate in electronic limit order books, the passive orders are usually not analyzed or even available for analysis. In Figure 2.1 a limit bid and a limit ask order become active and enter the order book.

The limit orders can be updated or cancelled at anytime until they are executed (i.e. the trade commences) or they expire due to the cancellation provision specified in the order. These provisions are usually time related. For example Good Till Close (GTC) limit orders are automatically cancelled at the end of the daily continuous trading. The exact content of these are specified in the market model of the exchange and can include provisions such as Fill-or-Kill (FOK) and All-or-None (AON). The limit order update semantics vary, but the in general it can be perfectly replicated with cancelling the old order and posting a new one at the same time.

Limit orders can be posted as far from the current best price as desired. They can only be executed at the defined limit price (or better), but this also means that the *time-to-execution* is uncertain and they might never be executed. Limit orders are also subject to *adverse selection*, since they are most likely to execute when the market moves against them (Menkveld and Zoican 2014).

2.3 Market orders

Additionally to the limit orders the market participants can submit *market orders*, which remove or *take liquidity* from the market. Market orders are matched against

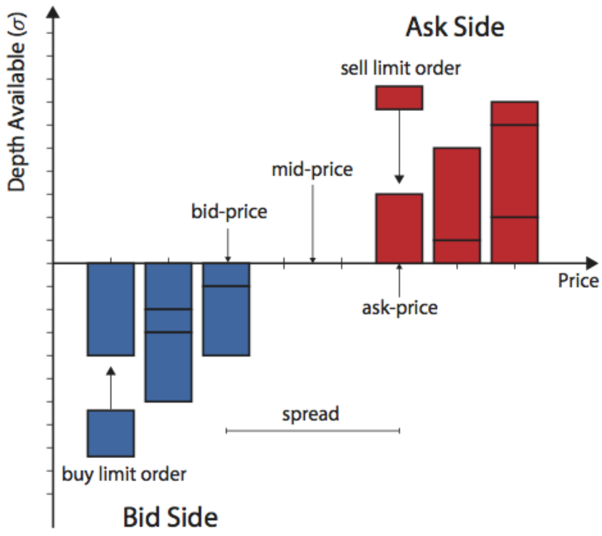


Figure 2.1 *Basic model of a limit order book (Gould et al. 2013)*

the active orders in the LOB and they are what ultimately lead to limit order execution. Two market orders of opposite sides can not match with each other since the execution price would in that case be ambiguous.

Market orders consist of pure market orders which specify just the desired quantity to traded and the side, as well as limit orders price of which is generous enough to produce an immediate match. It's practical to treat both pure market orders and spread crossing limit orders as *market orders*. This is essential for the empirical analysis conducted, since the two types can not be individually identified from the data set used. (Gould et al. 2013, pp. 2 – 4)

Provided that there is sufficient liquidity in the other side of the book market orders are immediately executed at the best available price. In the case of pure market orders, if the full order quantity is not available at the best bid (ask) the remainder is cancelled. In the case of spread crossing limit orders, the remaining amount enters the book as a active limit order with the corresponding limit price. For the purposes of analysis this event can be treated as 2 discrete order submissions:

- bid (ask) market order of depleting the sell (buy) side best price queue
- new bid (ask) limit order of the outstanding quantity at the corresponding price

There is no fundamental difference between limit and market orders in many markets. The presence of pure market orders is not required for the function of a limit order book market, because they can be replicated by the use of limit orders. Not taking the latency into account, market orders guarantee an immediate (and certain) execution, while having an uncertain execution price. When posting an market order, the participant *pays the spread*, which increases the *transaction costs* accrued. (Gould et al. 2013)

2.4 Limit order book state

A single valid order book state ($\mathcal{L}$) after matching represents an equilibrium in a sense that all the trades possible have been executed and no trades can occur without new orders or order updates entering the book. Order cancels can also impact the state but they should not lead to a new match and thus, a trade. Nevertheless, formally every new *order book event* e leads to new $\mathcal{L}_{i+1}$, which is function of the previous state $\mathcal{L}_i$ and the new order book event e_i.

These states are immediately distributed to DMA market participants after the matching engine has executed all the possible trades. See Figure 2.2 for an visualized example for a LOB state.

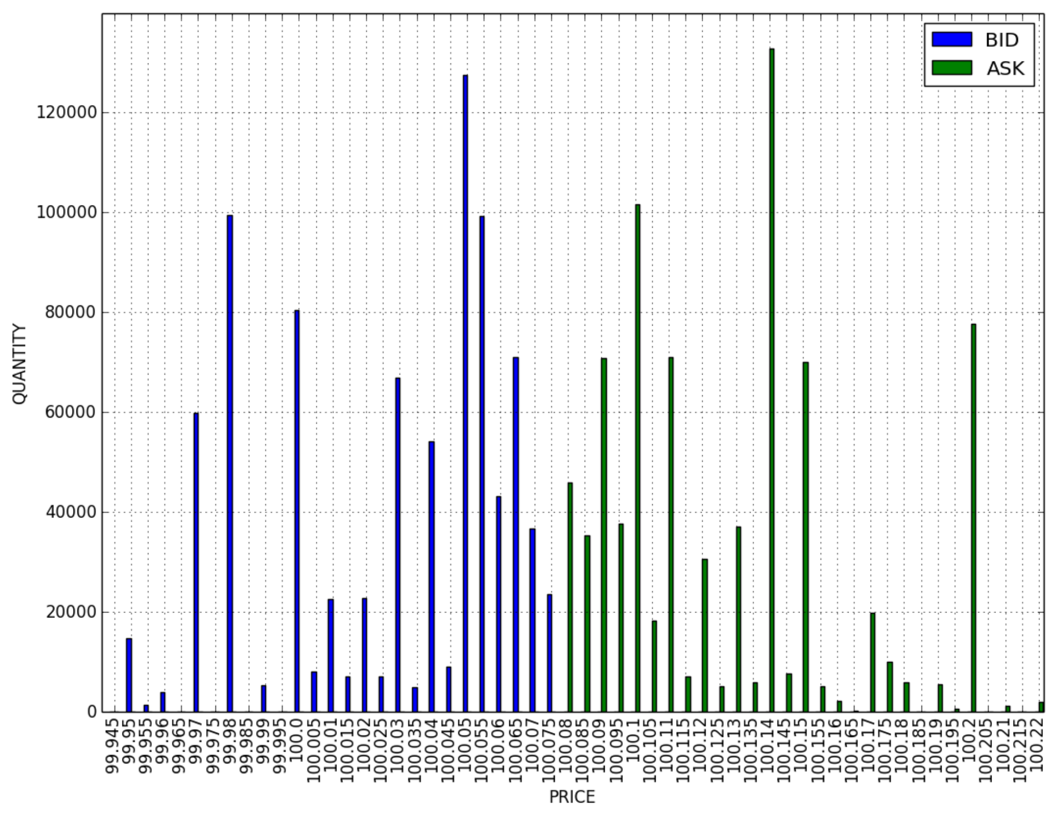


Figure *2.2 Example state $\mathcal{L}$ of an order book distributed to market participants*

Under these assumptions it's appropriate to define basic variables that represent elements of the LOB at the time t prevailing order book state $\mathcal{L}(t)$. These variables include tick size (δ), lot size (σ) and the spread ($s(t)$). This paper mainly uses the mathematical semantics defined previously by Gould et al. (2013).

Tick size

The tick size (δ) is not strictly an element of a single order book state but usually it's defined for a longer period. It represents the minimum price movement of a trading instrument and it's generally derived from the features of the underlaying (e.g. the company in the case of common stock). The tick size is additionally the minimum bid price. Other factors include the daily trading volume and other various exchange rules.

The tick size is used to emphasize the time priority of limit orders encourage early limit order insertions. If the tick size is very small or even non-existent, the market participants could gain priority and overtake existing orders in the queue by posting order with just barely better price.

In some exchanges (such as OMX Nordic), the tick size depends on the attributes of the *order*, not just of the *order book*. In such cases the tick size is a *order level* consideration. A rough estimation of OMX Nordic tick size derivation procedure is given in the Chapter 4. Previous studies have generally assumed a fixed tick size. The incongruities arising from this fact are taken into account in the corresponding analyses.

Lot size

Lot size (σ) is the smallest amount of underlaying securities that can be traded in one transaction. This naturally sets the minimum size of an order. Lot sizes are usually set rather informally for stocks, where the more expensive stocks trade with lot sizes closer to 1.

Some exchanges do allow orders with sizes less than the set lot size. These are called *odd lot orders* and they are matched in a separate order book from the *even orders*. These order books are generally have less traffic and have more stringent rules to reduce the risk of adverse selection. This was the case in the OMX Nordic during the time period of the data set and the analysis is focused on the even lot orders.

Since there is essentially no cost to produce matches in electronic order books, the lot sizes have decreased and in many markets they have been all but abolished with all securities trading with lot sizes of 1. Higher lot sizes encourages the placement of larger limit orders and reduces the possibility of institutional trades to split large orders to smaller ones to reduce market impact. (Gould et al. 2013, pp. 10 – 11)

Spread

Level 1 order book is the combination of current highest bid $b(t)$ and lowest ask $a(t)$ in order book state $\mathcal{L}(t)$ and the respective available quantities. In other words, incoming market orders would be executed against these prices. The level 1 LOB is usually given in the form of $b(t) - a(t)$. For example: €100 – €101.

The difference between these two prices is the *spread*:

$$S(t) = a(t) - b(t), \tag{2.1}$$

which is the amount market order submitter would have to pay gain immediacy compared to the limit order at the best price. The absolute spread in the earlier case would be €1, but for the purposes of this paper it's by and large more convenient to define it based on ticks

$$s(t) = \frac{a(t) - b(t)}{\delta} \tag{2.2}$$

In the case of €1 spread with a tick size of $\delta = e\,0.25$ the spread could be also defined as 4 ticks. (Farmer et al. 2005)

The size of bid-ask spread can be considered to be a measure of liquidity, if liquidity is measured as "the cost of turning round a position over a short period of time" (Kyle 1985). Lower market spread sizes also mean that the market values the immediacy and certainty of market orders versus the uncertainty of limit orders. (Gould et al. 2013, p. 5)

Midpoint Price

Apart from the inside spread, the current price of an instrument is defined as the mean of the best bid and ask. This is called the midpoint price or mid price $m(t)$:

$$m(t) = \frac{b(t) + a(t)}{2} \tag{2.3}$$

In the previous example case the $m(t)$ would be €100.5. Note that if the absolute spread in ticks is of an odd quantity, $m(t)$ is not a price with which new limit orders can be submitted. Midpoint are sometimes considered to be the "fair price" of the moment for the corresponding equity. At the $m(t)$ both trade participants would share equal amounts of the transaction costs associated with the spread. As such, midpoint price is the regularly quoted by the exchanges.

Relative Prices

Relative prices of incoming orders are defined as a function of the prevailing order book state. The common use case in literature is the defining incoming prices as the distance from the best asks (bids) on the corresponding side of the limit order book. The distance is the available price levels between the P_l or $b(t)$ or $a(t)$. Formally, we can define relative price of limit buy order with price P_l to be

$$\Delta = \frac{b(t) - P_l}{\delta} \tag{2.4}$$

and similarly for the sell order

$$\Delta = \frac{P_l - a(t)}{\delta} \tag{2.5}$$

where the δ is the stock's tick size at the time t. Note that the relative prices are positive when the price of the incoming order is worse that the available best prices. If the limit buy (sell) order arrives at the best bid (ask), the relative price distance is 0. Orders that have an immediate price impact and hit the book within the spread have a negative relative price.

The relative prices can also be defined as the distance from opposite quote, which is guaranteed to have a positive sign for all passive limit orders. Again, defined for the bids:

$$\Delta_o = \frac{a(t) - P_l}{\delta} \tag{2.6}$$

and for the asks:

$$\Delta_o = \frac{P_l - b(t)}{\delta} \tag{2.7}$$

This kind of relative prices are used in theoretical limit order book dynamics, since they capture the difference from using a market order regardless of the spread. The probability of limit order insertion is thus modeled as a function from the opposite queue, not the current best price at same side.

2.5 Empirical studies

An increasing amount of data is available on limit order book market microstructure. Empirical studies of limit order books usually rely on large time series with differing levels of access to variables governing the order book states. Most detailed data sets have access to every limit and market order insertion, update and cancellation, while some have only the trades and quotes, such as the popular and freely available NYSE TAQ (Trades and Quotes) database. Some of the well established statistical behavior can be deduced from these level 1 datasets, while some order level stylized facts rely on more detailed time series.

The time series can also be defined on a basis of their detail and ”snapshot frequency”. Gould et al. (2013, p. 4) uses the sampling procedure to define three categories of limit order book time series $t_1, ..., t_n$:

- t_i regularly placed in time with τ seconds between them – *τ-second timescale*
- t_i corresponding to arrivals of orders and cancellations – *event-by-event timescale*
- t_i corresponding to trades – *trade-by-trade timescale*

Note that this listing is not exhaustive and can be extended e.g. by defining own category Order-by-Order for datasets have only order insertions, but no other types of events such as cancellations. In a similar fashion, Cont (2011, p. 2) uses the frequency of limit order book snapshots to define rough categories for limit order book data sets and their usage in financial analysis:

Table 2.1 *Time series classification*

Category	Time scale	Usage
Ultra High Frequency (UHF)	$10^{-3} - 0.1$ s	Microstructure
High Frequency (HF)	$1 - 100$ s	Trade execution
Daily	$10^3 - 10^4$ s	Trading strategies

For the purposes of studying the order flow characteristics, the availability od *event-by-event timescale* or *Ultra High Frequency* time series is essential. Data sets with a lower detail level can be used to verify previous statistical observations, but the studies usually rely on number of generalizations and estimations, which may affect the results. The event-by-event nature of data sets such as TotalView ITCH Feed allows the tracking of individual orders throughout their lifetime as passive orders in the limit order book and can provide quantitative frameworks for short term market variables. (Cont 2011)

Problematic for the most detailed data sets is the amount of computational power and storage space required to conduct exchange wide analysis throughout a long time series. Some trading strategies place and cancel orders in rapid bursts, which further bloats the time series and makes the $\mathcal{L}(t)$ more unstable and more complex.

For the purposes of this thesis, the empirical studies presented in Table 2.2 have been significantly influential in their estimation methods. The list consists of both purely econometric approaches (i.e. try to establish or test stylized facts) and studies that leverage and test theoretical LOB dynamics. The studies have also analyzed the real life implications of the reported values, upon which this thesis builds a significant amount of its conclusions. Note that most of these studies build upon high or ultra high frequency data sets. This choice was deliberate, since the methods used to analyze such time series are comparable to the methods used in this thesis. The categorization presented in Table 2.2 simplified, since the detail might be limited by other factors, such as the number of relative price levels available.

Table 2.2 *Empirical studies with similar methods and samples*

Reference	Exchange	Sampling method	Frequency
Bouchaud et al. (2002)	Paris Bourse	Order-by-Order	UHF
Farmer et al. (2005)	Paris Bourse	Trade-by-Trade	HF
Cont et al. (2010)	Tokyo Stock Exchange	Order-by-Order	UHF
Cont et al. (2014)	New York Stock Exchance	Order-by-Order	UHF
Gu et al. (2008)	Shenzhen Stock Exchange	Event-by-Event	UHF
Farmer et al. (2005)	London Stock Exchange	Event-by-Event	UHF
Zovko and Farmer (2002)	London Stock Exchange	Event-by-Event	UHF
Toke (2011)	Euronext Paris	Order-by-Order	UHF
Zhao (2010)	International Petroleum Exchange	Order-by-Order	UHF
Hautsch and Huang (2011)	NASDAQ	Event-by-Event	UHF
Hollifield et al. (2004)	Stockholm Stock Exchange	Order-by-Order	HF

2.6 Theoretical models

The models for limit order book dynamics can be divided into two main categories, *zero intelligence* and *perfect rationality* (also know as static and dynamic equilibrium models) (Zovko and Farmer 2002). The models make differing assumptions of the agents posting orders and their behaviour. The perfect rationality models have a lot common with the traditional academic economic literature where as zero intelligence models apply the methods of *econophysics*.

The zero intelligence models in continuous time are of a greater interest to this paper, but a short overview and history of equilibrium, as well as discrete time zero intelligence, are provided. Zero intelligence models use point processes to reflect discrete order book states. Point process is a random process, where the states (or points) are isolated from each other. Stochastic Markovian (i.e. memoryless) queues with arrivals occurring according to some statistical distribution can be seen as a general model of a limit order book.

2.6.1 Perfect Rationality

Perfect rationality models present limit order book as a market, where agents post and cancel orders driven by new information. These agents are assumed to be perfectly rational. In other words investors choose portfolio strategies perfectly in order to maximize their personal utility. This approach usually assumes that the investors can buy or sell assets frictionlessly. It is complicated by the fact that the investors can not be sure of limit order execution and simple portfolio of holdings is extended by a portfolio of limit and market orders. (Parlour and Seppi 2008, pp. 4 – 5)

Many early perfect rationality models (e.g. Kyle (1985) and Glosten (1994)) make assumptions of trading motivations and strategies of market agents dividing them to *informed* and *uninformed* traders. The informed traders are assumed to know the *fundamental* or *true* value of the underlying security. This assumption leads to a market where perfectly rational informed traders use market orders to pick-off uninformed traders once they post limit orders with more generous prices than the true fundamental value.

Parlour (1998) and Foucault (1999) presented dynamic equilibrium models for limit order books, where the market model is considerably simpler than real life LOBs. Nevertheless they were able to explain some strategies observed in real life LOBs, such as: wider spread leads to greater proportion of limit orders, thick order books lead to more market orders in order to gain priority and volatility effect on limit price setting.

Work by Rosu (2009) and (2014) extended these models and was able to provide and explanation for multiple empirically observed limit order book phenomena. Additionally, they provide model for informed investors to prefer split limit orders in certain situations, for example to minimize their market impact.

A problem for the dynamic equilibrium models, is that their predictions are often times challenging or impossible test with the available limit order book data. Due to the great sizes and number of observations in data sets such as in this paper, the actual motivations of traders, their personal utility or even the information content is hard to assess in a robust way. Nevertheless, they present attractive explanations why stylized facts arise from rational behavior by heterogeneous market participants.

2.6.2 Zero Intelligence

Whereas perfect rationality approaches assume order flow, and thus $\mathcal{L}(t)$, to depend, at least in part, on exogenous variables, zero intelligence models order flow as a set of stochastic processes. The processes define the arrival rates of limit orders, market orders and order cancellations. The processes can be calibrated using the wealth of data available on limit order book events.

The first real zero intelligence LOB model was introduced by Bak et al. (1997). It modelled limit order book state as a stochastic mass particle behavior. The particles moved across the price lattice following a random walk. While simple, the model was capable of reproducing various limit order book phenomena such as hump shaped average order book depths across price levels. More realistic zero intelligence models incorporating complex behavior have been introduced by Maslov (2000) and Challet and Stinchcombe (2001). These processes were defined in discrete time and assumed one or more traders arriving in the market at each discrete time step and placing a new order (i.e. a particle). These particles were allowed to evaporate exogenously and independently in a discrete state space. (Gould et al. 2013, pp. 25 – 26)

Continuous time zero intelligence models have been proposed by Daniels, Smith and Cont et al. (2010). The early models assumed tick size to approach zero ($\delta \to 0$) while Cont et al. defined a model where the orders arrived at discrete relative prices. Arrival rates of limit orders at every relative price was assumed to be a independent Poisson process. Similarly the inter-arrival times of market orders were assumed to be exponentially distributed. Cont et al. also proposed a method to estimate the parameters for the model from lower depth limit order book historical data, if full order level data was not available.

The model by Cont et al. has been later empirically criticized by Toke (2011). The independence and exponential distributions of inter-arrival times were found not to appear in real life limit order book. Instead, the order book event arrivals were found to be clustered and have a self-exciting property. Toke called this effect ”market making” in order book. The effect was modeled with 3 mutually exciting behaviour MM, LL and LM:

- MM - market orders increasing market order arrival
- LL - limit orders increasing limit order arrival
- LM - market orders increasing limit order arrival

These effects were modelled with Hawkes processes, which are point processes with a time varying intensity parameter. Toke found this model fit the empirical data better than the homogeneous Poisson arrivals of Cont et al. (2010) by leveraging all three ”market making” effects. Additional flaws of zero intelligence models were pointed out by Smid (2015, pp. 6 – 8). He proposes a Generalized Zero Intelligence model (GZI), which attempts to address the following issues:

- all orders are not of unit size
- agents do not behave in a completely random fashion

The use of unit size orders is problematic especially in the case of market orders, where the orders can not be assumed to be comprised of several one-lot orders without violating the Poisson assumptions. The second item is also related to the Poisson arrivals, since the clustering of order book events can be also thought as market participants responding strategically to changes in order flow.

Typically the zero intelligence models are able to produce easily testable hypotheses. These include factors such as probability of a price increase (a *uptick*) and probability of executing before change in $m(t)$. Traditional Poisson zero intelligence models assumptions have nevertheless been shown to not hold in real life limit order books. Exogenous effects such as information arrival have not been explicitly incorporated to zero intelligence models and they perform poorly at a longer time scale compared to the dynamic equilibrium models.

This thesis focuses on the limit order book model proposed by Cont et al., but criticisms by Toke are additionally explored. The actual assumptions of order flow required by these models are further discussed in the next chapter and their empirical ramifications are explored in the Chapter 5.

3. LIMIT ORDER BOOK EVENTS

Agents posting orders on the exchange must choose between limit or market orders. Market orders are guaranteed immediate execution while the price remains uncertain. Regular limit orders do not have the feature of immediate execution but the price is settable beforehand and the agent does not have to cross the spread. These order book events combined with cancellations define the set of order book events.

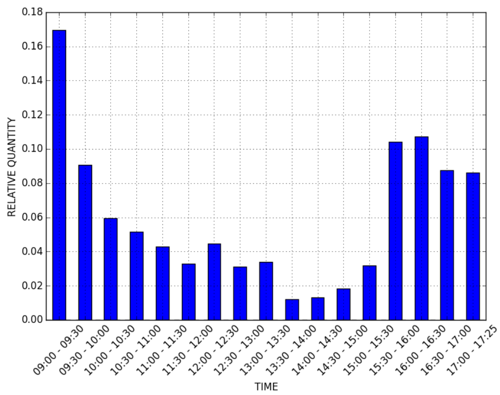
For these events, a multitude of stylized facts and model assumptions have been defined. In the field of econophysics, stylized facts are interpreted as "facts recorded by statisticians" and "they concentrate on broad tendencies, ignoring individual detail" (Kaldor 1957). This chapter explores these facts in light of academic literature and defines them in more detail for empirical testing in Chapter 5.

3.1 Seasonality

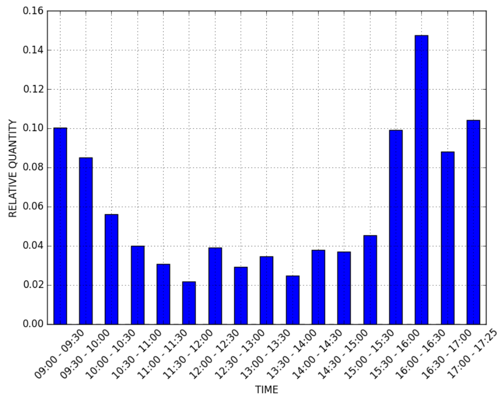
Arrival times of limit orders during the continuous trading exhibit significant seasonality and correlation. (Lorenz and Osterrieder 2009, p. 7)(Bisiere and Kamionka 2000, p. 44) The traded volume regularly follows an *u-shaped pattern*, where the volume is highest just after the market open and just before the market close. Similar pattern can be observed in the arrival activity of limit orders. See Figures 3.1a and 3.1b for an example.

Different order book events have individual diurnal seasonality profiles. Cancellations and trades (market orders) have similar seasonality characteristics with traded volume, but they can at least partly be attributed to limit order seasonality through the self-exciting properties of limit order submission present in recent stochastic limit order book models.

Seasonality has ramifications for general limit order book study and some studies have detrended the data before analysis. This is usually done by splitting the trading day into periods and diving or subtracting them by the mean.



(a) NOK1V traded amounts per half hour on September 2nd 2010. Last period has been scaled accordingly to factor for the 5 minute length difference



(b) NOK1V limit order arrivals per half hour on September 2nd 2010. Last period has been scaled accordingly to factor for the 5 minute length difference

3.2 Arrival times

The times between discrete order book events (limit orders L, cancellations C, and trades T) are the inter-arrival times $\Delta t(L, C, T)$. The are usually called *durations* and are estimated by the use of *duration processes* in LOB modeling. It's usually defined independently for every single event type, such as limit orders:

$$\Delta t(L_n) = \begin{cases} t(L_n) - t(L_{n-1}) & \text{if } n \geq 1 \\ t(L_1) & \text{if } n = 0, \end{cases} \tag{3.1}$$

where L_n is the nth limit order for a particular stock. The duration can be similarly defined for any type of limit order book event or subclasses of the events. The aggregate inter-arrival times define the *arrival rate* for a given time period.

According to a empirical study by Lorenz and Osterrieder (2009, pp. 7 – 8), inter-arrival times of orders are all in all Weibull(k, λ) distributed. These are the arrival times of orders of any side, quantity or price. The duration process for such case is

$$f(\Delta t; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{\Delta t}{\lambda}\right)^{k-1} e^{-(\Delta t/\lambda)^k} & \text{if } \Delta t \geq 0, \\ 0 & \text{if } \Delta t < 0, \end{cases} \tag{3.2}$$

where $k > 0$ is a *shape parameter* and $\lambda > 0$ is a *scale parameter*. The Weibull distribution is reduced with $k = 1$ to an exponential distribution (Rinne 2008):

$$f(x, \lambda) = \lambda e^{-\lambda x} H(x), \tag{3.3}$$

where $H(x)$ is the Heaviside step function.

The exponential distribution as a limit order inter-arrival model has been proposed by Cont et al. (2010). It should be noted that the empirical study concerning Weibull distribution did not take account the relative price of an order. For a relative price level Δ we can define individual inter-arrival time $t(L_n, \Delta) = t(L_n, \Delta) - t(L_n, \Delta)$ and inter-arrival time exponent $\lambda(\Delta)$. Another distributions suggested to be used as the relative price limit order duration process include gamma distributions and power laws. (Gould et al. 2013, pp. 27 – 28; Cont et al. 2010)

These inter-arrival times are a important factor in mathematical modeling of limit order books and their dynamics. Many models use Poisson processes to model order arrivals, which requires inter-arrivals to be exponential distributions and independent from each other. For example, Cont et al. (2010, pp. 550 – 551) uses independent Poisson processes to model individual inter-arrival times of order book events. Denoting the distance from the **opposite** side best quote Δ_o, he defines the following variables:

- Limit orders arrive at independent exponential rates with rate $\lambda(\Delta_o)$
- Market orders arrive at independent exponential rates with rate μ
- Cancellations arrive at a rate proportional to the number of active limit orders at a given level

Δ_o has the convenient characteristic that it is always greater than 0. Limit order with $\Delta_o \geq 0$ would qualify the incoming order for a immediate match and would thus be regarded as a market order. For the most liquid stocks that trade predominately with minimum spread of 1 tick the opposite side distance corresponds with the regular price with $\Delta = \Delta_o - s(t) = \Delta_o - 1$.

3.3 Arrival position

Limit order's relative price at the moment of order insertion as an active order can be called the *arrival position* (Δ) of the order. Similar measure can be defined for cancellations as the relative price of the cancelled order at the time of the cancellations. The cumulative relative position of incoming limit orders have been previously empirically found to have power law tails by Potters and Bouchaud (2003) in the Paris Bourse and by Zovko and Farmer (2002) in the London Stock Exchange. In general form, continuous power laws are determined by the decay

$$p(x) \propto f(x)^{-\alpha}, \tag{3.4}$$

where α is the exponent of the power law. Various distributions are proportional to this decay and all of them are power laws. Power laws describing distribution tails usually additionally incorporate a scaling factor and tail limits x_{min} and x_{max}. This is the period, for which the power law is fitted. The ask side placement has essentially no limits in order placement, while the bid side relative prices of orders are additionally limited by the tick size price floor δ.

Using the power law function defined by (Potters and Bouchaud 2003, p. 134) and denoting the distance from the best price in the respective side of the limit order book as defined by the Equations 2.4 and 2.5, the cumulative probability of order of any size arriving at that position is

$$P(\Delta) \propto \frac{\Delta_0^{\mu}}{(1+\Delta)^{1+\mu}}, \tag{3.5}$$

where μ is the constant exponent, also called the decay.

For three stocks in Paris Bourse the best fit was acquired with $\mu = 0.6$ (Potters and Bouchaud 2003, p. 134) and in London Stock Exchange with $\mu = 1.5$ (Zovko and Farmer 2002, p. 2). One explanation offered for this difference was the fact that all order flow in Paris Bourse was electronic while this was not the case for the London Stock Exchange. For Paris the fit was observed with $x_{min} = 10$ and $x_{max} = 1000$, depending on the stock. In LSE, the tail was fitted in period of $[10, 2000]$ ticks, but the authors did not have confidence in the estimates for lower arrival rate stocks. Beyond x_{max}, an exponential decay overtook the power law as fewer orders are placed further from the best prices..

As mentioned, power law tails of arrival positions have been constantly observed in multiple studies done in different limit order markets and it has achieved a status as a stylized fact. The exponents presented have generally been defined exchange wide with low variance from stock to stock, with some differences between low and high liquidity stocks.

Some theoretical models use this stylized fact when estimating the order arrival rates from data sets of lesser detail. Cont et al. (2010) use the power law estimation to derive arrival rates for other order book levels from NASDAQ Trades and Quotes (TAQ) data, which directly allows to estimate values for the best price levels. If the exponents of the power law decay are known, the arrival rates for top levels of the order books can be used to deduce the arrival rates for further price levels.

The general use of power laws in explaining limit order book phenomena have nevertheless been criticised by Clauset et al. (2009). According to the authors the use of continuous distribution leads to a questionable results due to discrete nature of the empirical data. In any case, the authors stress the need to define x_{min} formally using appropriate estimators. For a detailed discussion on the use of power laws in empirical data see Alstott et al. (2014) and Clauset et al. (2009)

3.4 Price Impact

Price and market impact are the changes in the observed prices $b(t)$, $a(t)$ and $m(t)$ caused by a single event or a set of events. These are important considerations for e.g. traders wanting to transact large quantities of securities in a short time span. The price change is usually defined as the change in $m(t)$, although the changes in $b(t)$ or $a(t)$ are usually more important for individual market participants, depending on the direction he wants to trade.

3.4.1 Instantaneous and permanent components

In a study by (Gould et al. 2013, pp. 17 – 18), the total price impact of an order book event e has been divided into two distinct components:

- *instantaneous impact* is the immediate price change caused by event e hitting the order book and directly affecting $\mathcal{L}(t+1)$.
- *permanent impact* is the change in future order flows caused by event e hitting the order book and indirectly affecting $\mathcal{L}(t+n)$, where $n \in \mathbb{N}$.

The permanent impact is impossible to exactly quantify from empirical data, since the random fluctuations and seasonality of order flows can't be isolated from order flows caused by the event e. Thus the quantification would require comparing situations, where the event e happens and where it does not. This is appropriate for studying using simulations, but for a Zero Intelligence model to account for any kind of permanent impact, the model must incorporate some kind of self-excitement or regime shift methodology. For a instantaneous impact a formal definition can be given as

$$\Delta m(t)_e = m(t^{'}) - \lim_{t \to t^{'}} m(t), \tag{3.6}$$

where $t^{'}$ is the time of the order book event. For empirical purposes, it is purposeful to define price impact in fixed time scale basis. Let t_i, be the ith observed starting time of time span in the sample. For every aggregated collection of events in the time span $[t_{i-1}, t_i]$, the total price impact measured is the price change in the following time span $[t_i, t_{i+1}]$

$$\Delta m(t_i) = m(t_i) - m(t_{i-1}), \tag{3.7}$$

which is easily determinable from empirical order book data. While this approach does not allow directly to study price impacts of individual order book events, it has various real world applications. This kind of aggregated impact has been previously been studied how it relates with the imbalance of *trades* executed against either sides of the order book.

For the purposes of this paper, the Order Flow Imbalance introduced by Cont et al. (2014) is used, which combines the three main types of order book events: order insertions, cancellations and trades (market orders). This makes it perfectly reproducible by the data provided by the TotalView ITCH feed. The following Section explores the OFI model in more detail.

3.4.2 Order Flow Imbalance

During the continuous trading, there exists two queues, the bid and ask side, properties of which are affected by the order book events. Some of the events add *queuing nodes* to the queue and others remove them from the queue. Limit orders can be modeled as new nodes entering the queue and cancellations as nodes leaving the queue before being served. Trades correspond with queuing nodes being served with market orders and then accordingly leaving the queue. In a simplified, stylized order book the midpoint price $m(t)$ and available prices for immediate execution, $b(t)$ and $a(t)$, are purely defined by these two queues.

This means that the order flow accrued during the time span $[0, ..., t]$ perfectly defines the $\mathcal{L}(t)$. Thus it's clear that limit incoming order book events themselves can have an impact on the midpoint price $m(t+1)$. This is most apparent in situations where an order or it's cancellation directly affect either of the best prices $a(t+1)$ or $b(t+1)$. Since the short term midpoint price movement is purely defined by the level 1 order book, it's feasible to define it in terms of Order Flow Imbalance (OFI) at the best bid and ask.

In real order books, the queues at other price levels do affect the price movement, but lets first considerer stylized order book, where orders only arrive at best price levels. For the stylized model of the order book, where the order insertions and deletions and cancellations happen only at the best bid and all of the price levels have the depth D, Cont et al. (2014) defines the bid side OFI at the time t_k and market impact as

$$\mathrm{OFI}_{b,k} = L_{b,k} - C_{b,k} - M_{s,k} \tag{3.8}$$

$$\Delta_{b,k} = \mathrm{OFI}_{b,k} / D, \tag{3.9}$$

and for the ask side

$$\mathrm{OFI}_{s,k} = -(L_{s,k} - C_{s,k} - M_{b,k}) \tag{3.10}$$

$$\Delta_{s,k} = \mathrm{OFI}_{s,k} / D, \tag{3.11}$$

where L denotes limit order insertion, C cancellation and M market order quantities at the best price levels of the LOB. Net bid side order flow's sign is considered to be positive, because increasing amounts of nodes queuing in the bid side is adding pressure for the price to increase. Combining these with the Equation 2.3 yields the general midpoint price impact ΔP_k

$$\Delta P_k = \frac{\mathrm{OFI}_{b,k} + \mathrm{OFI}_{s,k}}{2D} + \epsilon, \tag{3.12}$$

where ϵ is the truncation error. This model holds just in a statistical sense, since the assumptions of the stylized order book do not generally hold for real life limit order books – for example, events do not just occur at the best price levels. Cont et al. (2014, p. 7) thus proposes a noisier relation with price changes and the OFI:

$$\Delta P_{k,i} = \beta_i \, \mathrm{OFI}_{k,i} + \epsilon_{k,i}, \tag{3.13}$$

where slope β_i is the *price impact coefficient*. The coefficient is a measure of the price movement relative to the OFI. Higher values imply higher short term volatility since lower imbalances can have greater effects on the price. The equation also assumes a linear correlation with $OFI_{k,i}$ and $\Delta P_{k,i}$. For every time period k, the coefficient values can be estimated with the use of simple regression.

Another measures of short term price volatility include estimators such as *Order Book Slope* and *Trade Imbalance*, but since the main focus of this paper is order flow and its dynamics, Order Flow Imbalance is the chosen measurement. Moreover, the trade imbalance was found to be a poorer predictor of price movement than order flow imbalance by Cont et al. (2014).

4. NASDAQ OMX NORDIC

NASDAQ Nordic OMX runs security exchanges in Helsinki, Copenhagen, Stockholm, Iceland, Tallinn, Riga and Vilnius. These market have differing market models and this chapter is limited in scope to Nordic exchanges in Helsinki, Stockholm and Copenhagen. The times are given in Central European Time (UTC +1), which is the local timezone for Stockholm and Copenhagen. Helsinki is one hour ahead with Eastern European Time (UTC +2).

The market model described here conforms to the version 2.2 released in April 1st, 2011 (NASDAQ 2011). It has been since updated with further revisions, but due to the data set's time period, a now outdated version is used. It should be noted that for the period of August 2010 – March 2011 the market model used the version 1.1 (NASDAQ 2010) of the document, which had a differing tick size tables. The differences between market models are taken into account in the further analysis.

4.1 Market model

The three equity exchanges are run as purely electronic limit order book markets with three distinctive stages during the day (Table 4.1). This thesis focuses on the continuous trading, which lasts for 8 hours and 25 minutes in Helsinki and Stockholm. In Copenhagen this period lasts for 7 hours and 55 minutes.

Continuous trading is preceded by a pre-open call and an uncross. A similar auction and an uncross is performed after the close. The opening and closing auctions are used to increase liquidity and reduce adverse selection in stocks with fewer trades by conducting all the possible trades with such price P that total volume Q_P satisfies

$$P = \arg\max_{P} Q_P \tag{4.1}$$

Because the focus of this study is in the characteristics of the continuous order flow exhibited during the regular trading hours the auction and uncross dynamics or procedure are not defined here in a great detail. Also, the limit order book does not have a traditional valid state during the auction periods and it's possible for bid and ask queues to cross (i.e. such state that $b(t) \geq a(t)$). This means that variables such as relative prices or spread are not consistent compared to the ones defined for the continuous trading period.

Table 4.1 *The regular equity trading hours*

	Stockholm	Helsinki	Copenhagen
Pre-open	08:00	08:00	08:00
Uncross	09:00	09:00	09:00
Continious Trading	09:00 – 17:25	09:00 – 17:25	09:00-16:55
Pre-close	17:25	17:25	16:55
Uncross	17:30	17:30	17:00
Post-trade	17:30	17:30	17:00

The exchanges accept limit and market orders during the continuous trading according to principles set in Chapter 2. The matching follows price-internal-display-time or price-display-time priority. The priority defines the order of nodes in the bid and ask queues. In this context, the word internal refers to market participants (i.e. stock brokers). This allows trade internalization within participant's clients. For the purposes of this paper, the microstructure of individual price queues is not essential.

4.1.1 Order attributes

Market agents can additionally set various attributes to their orders. These include reserve, pegged and hidden orders. These can not be individually detected from the TotalView ITCH feed. Pegged limit orders can be set as an offset from current $a(t)$, $b(t)$ or $m(t)$. The orders are updated if the variable they are pegged onto is changed. This is announced on the feed as two distinct events:

- Complete cancellation of the old order.
- Insertion of a new order with recalculated price. Other order semantics, such as quantity, remain unchanged.

Reserve (or iceberg) limit orders consist of a displayed and a hidden portion. When the displayed portion is completely depleted, a new display order is inserted into the order book with a new time priority. This order is of equal size to the original displayed portion and the non-displayed part quantity is decremented accordingly. This is advertised on the feed as a regular new limit order submission propose no problem for analysis.

In stylized limit order market, outstanding orders remained active in the book until they were explicitly cancelled. In order to reduce probability of adverse selection, NASDAQ provides participants with various types of Time in Force attributes, that limit the possible lifetimes of orders. Orders in OMX Nordic markets can have 5 types of Time in Force attributes:

- Immediate-or-cancel (IOC)
- Good-till-market close
- Good-till-cancelled (GTC)
- Good-till-time (GTT)
- Day order

Detailed differences between these order attributes are beyond the scope of this thesis. All market orders are assumed to execute immediately. Limit orders are valid until cancelled, executed or the market close, whichever comes first. Multi-day orders are not tracked since the technical representation of the order book starts empty every morning. All outstanding active orders are considered implicitly cancelled at the market close. Multi-day orders will be re-inserted and advertised on the feed at the next market open. Typically the outstanding orders are additionally valid for the closing auction and uncross, but since only the order flow during the continuous trading is studied, the cancellations can be implied to happen at market closing events.

Tick sizes

The three exchanges all have own tick size derivation tables which is natural since all use different quotation currencies. Additionally there during the data set there were two differing tick size regimes in force from June 2010 – March 2011 and April 2011 – May 2013.

In the literature, the tick size is often assumed to be constant for at least a single stock, but this is not the case in OMX Nordic. The Nordic Stock Exchanges enforce tick sizes at the limit order price level. This means that the required tick size δ_l for order l is not based on $m(t)$, $b(t)$ or $a(t)$ but on P_l. This means that even if the midpoint price of an security lies in a tick size interval $\delta = 0.01$, an order with precision of 0.005 is still valid if its price is low enough to lie in an appropriate tick size interval.

For this reason, Equations 2.4 and 2.5 can give unrealistic results when P_l and $b(t)$ or $a(t)$ are far enough away from each other. By using the simple definition of relative price as "the number of possible price levels between" we can calculate more precise values iteratively. This can also affect the calculation of spread, but such cases are assumed to be very rare and the $s(t)$ is always calculated using the tick size, in which $m(t)$ lies.

Trading halts and intraday auctions

The regular continuous trading hours can be interrupted due to a trading halt. The market model defines the detailed reasons for such halt, but in general they are imposed in anticipation of a news announcement or where there is a significant order imbalance. Trading can be also suspended by the regulatory authority. During the halt no new orders may be posted, but active orders can be cancelled.

Halts and resumptions of trading are announced in the ITCH feed with various stop codes defining the reason for the halt. Depending on the reason, the trading might be resumed with a new price discovery process (i.e. intraday auction) or the order books might be flushed. (NASDAQ 2011, pp. 21 – 22)

4.2 TotalView-ITCH data set

TotalView-ITCH is data feed available for certain markets run by NASDAQ OMX. It contains all order book and trading related messages and allows very precise tracking of individual orders with a millisecond precision. By the hierarchy provided by Cont (2011, p. 2) this can be considered to be Ultra-high frequency (UHF) financial data. Full order book state is thus reconstructible for every event that changes the state of the book (Hautsch and Huang 2011). The resulting time series are sampled on an *event-by-event timescale* as defined by Gould et al. (2013, p. 4).

The study is conducted on the complete TotalView-ITCH data from June 1st 2010 until May 31st 2013. During this period there were 783 trading days (i.e. days when there has been open order books at least in one of the exchanges). See Table 4.2 for a more detailed distribution. The length and detail level of the data set allows very precise testing of various order flow phenomena.

For the purposes of further analysis 4 stocks were selected from every Nordic exchange, 12 in total. Criteria for selection included factors such as sufficient liquidity, differing market capitalizations and industries. For cross-listed stocks, the exchange with the most liquidity was chosen. Similarly for the companies with multiple stock series issues the most liquid one was chosen.

Table 4.2 *Trading days in the data set*

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2010	–	–	–	–	–	22	22	22	22	21	22	23	154
2011	21	20	22	21	22	22	21	23	22	21	22	22	259
2012	22	21	22	21	23	21	22	23	20	23	22	21	261
2013	23	20	21	22	23	–	–	–	–	–	–	–	109

Due to the relative lack of daily liquidity securities with very small market capitalizations, as defined by NASDAQ OMX (2011), were left out of the consideration. See Table 4.3 for information of the stocks selected. The companies were also required to be listed for the full duration of the data set.

Table 4.3 *Securities chosen for further analysis*

Long Name	Currency	Exchange	Market Cap	Sector	Ticker symbol	ISIN
Sampo Oyj A	EUR	HEL	LARGE	Financials	SAMAS	FI0009003305
Nokia Oyj	EUR	HEL	LARGE	Technology	NOK1V	FI0009000681
Neste Oil Oyj	EUR	HEL	LARGE	Oil & Gas	NES1V	FI0009013296
Finnair Oyj	EUR	HEL	MID	Consumer Services	FIA1S	FI0009003230
Nordea Bank AB	SEK	STO	LARGE	Financials	NDA SEK	SE0000427361
TeliaSonera AB	SEK	STO	LARGE	Telecommunications	TLSN	SE0000667925
ABB Ltd	SEK	STO	LARGE	Industrials	ABB	CH0012221716
Active Biotech AB	SEK	STO	MID	Health Care	ACTI	SE0001137985
Danske Bank A/S	DKK	CPH	LARGE	Financials	DANSKE	DK0010274414
Vestas Wind Systems A/S	DKK	CPH	LARGE	Oil & Gas	VWS	DK0010268606
A.P. Møller - Mærsk B A/S	DKK	CPH	LARGE	Industrials	MAERSK B	DK0010244508
Bavarian Nordic A/S	DKK	CPH	MID	Health Care	BAVA	DK0015998017

ITCH feed documentation defines various types of *messages* that are sent for a corresponding order book events. Some events can trigger multiple messages, for example order cancellation that changes $b(t)$ and thus $m(t)$ would trigger order cancellations and re-insertions with an updated price for all corresponding pegged orders. This introduces unwanted noise to our data set, but the relative amount of pegged orders to regular orders is assumed to be small.

All continuous trading limit orders, trades and cancellations are extracted for these securities from ITCH feed messages. The total number of observations are listed in the Table 4.4. Note that market orders are not handled as a distinct unit, since in the data set they appear only as trades. ITCH feed order insertion messages are sent only for active orders in the order book, not for orders that are matched immediately. Hidden orders are naturally not advertised on the feed, but trades executed against hidden quantities are reported separately from regular trades.

As apparent from the Table 4.4, the number of cancels greatly dominate the total amount of trades especially for the stocks with the higher total order flow. Note that one limit order can be cancelled (or traded against) multiple times in case of partial fills or cancels. This is most evident when a sufficiently large market order

Table 4.4 *Total number of order, trade and cancel observations*

	Limit Orders	Limit Buy Orders	Limit Sell Orders	Trades	Cancels
NOK1V	194,236,767	96,748,052	97,488,715	9,501,563	184,735,204
NES1V	19,403,595	9,679,190	9,724,405	1,541,433	17,862,162
SAMAS	30,403,926	15,176,434	15,227,492	2,259,606	28,144,320
FIA1S	236,718	109,837	126,881	58,874	177,844
TLSN	107,983,477	54,157,501	53,825,976	4,063,777	103,919,700
ABB	69,525,633	34,864,729	34,660,904	2,184,132	67,341,501
ACTI	822,851	401,888	420,963	263,184	559,667
NDASEK	147,106,329	75,809,707	71,296,622	4,089,886	143,016,443
DANSKE	11,550,649	5,777,573	5,773,076	1,727,471	9,823,178
MAERSKB	19,183,399	9,490,058	9,693,341	1,694,505	17,488,894
VWS	30,199,507	15,011,689	15,187,818	3,115,637	27,083,870
BAVA	334,040	151,071	182,969	98,107	235,933
All	630,986,891	317,377,729	313,609,162	30,598,175	600,388,716

is submitted. It's likely to be filled by multiple active limit orders and generate a trade message for each of these matches.

Some authors using TotalView-ITCH or comparable data, use time based aggregation. For example Hautsch and Huang (2011) considered all trades executed within a half second with the same initiation variables to be part of a larger market order. The purpose of the study is not to examine the market impact caused by large orders and due to the clustered nature of order book events, time based aggregation is not used to deduce original orders from the order flow.

5. EMPIRICAL ANALYSIS

These order, trade and cancel messages combined with trade control messages (e.g. trading halts or mid trading auctions) were synthesized into complete time series limit order book states $\mathcal{L}(t)$. After that, features of order book events depending on the $\mathcal{L}(t)$, such as relative prices and price impact, were calculated.

The following Chapter 5 analyzes these two bodies of data and their interrelations in the light of the existing limit order book literature. To achieve this, the models presented in Chapter 3 are leveraged and explored further. The chapter is divided into three main parts: first part deals with the general overarching characteristics of order flow, such as seasonality and event sizes. Second part focuses on the observed clustering behaviour of order book events. The last section is dedicated for the Order Flow Imbalance approach to short term price impact of order book events. OFI is also used as a measure of order flow volatility across the time scales.

5.1 General characteristics

Apart from the absolute number of LOB events specified previously in the Table 4.4, many general distinctive order flow characteristics can be explored. These include the seasonality of the different types of flow, size of the events and relative position of the orders.

To facilitate more detailed examination, the events can be further divided into categories like the outline set out by Al-Suhaibani and Kryzanowski (2000, p. 1343 – 1345). Buy and sell limit orders are characterized by their relative price the moment they enter the LOB as active orders. Nevertheless, the order sizes are not used to categorize orders, unlike Al-Suhaibani and Kryzanowski (2000).

5.1.1 Relative prices and order sizes

To test the hypothesis that the cumulative relative prices of new orders follow a power law, Δ was calculated using the iterative versions of Equation 2.4 for buy and 2.5 for sell orders. The resulting distribution was fitted first stock wise to the power law in the form of Equation 3.4. To examine the exchange wide values, the distributions were aggregated as a volume weighted mean for a three exchanges and then fitted. The fitted estimates were acquired using a Python package (Alstott et al. 2014) implementing power law fitting procedure described in Clauset et al. (2009). The fitted values are shown in Table 5.1.

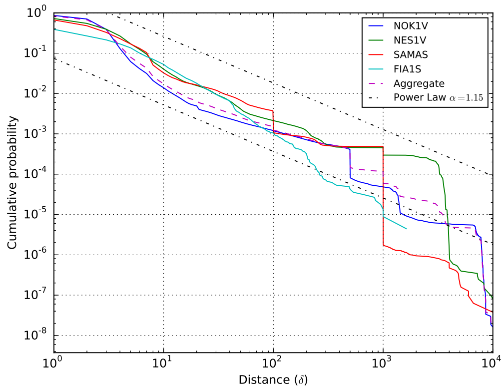
The package attempts to find the appropriate values for the exponent α and the lower bound x_{min}. To control for the effects of more rapid decay beyond the relative price $\Delta > 1000$, the x_{max} was set to 1000 ticks. This is similar to the observations made in London Stock Exchange and in Paris Bourse. The power law decay exponents for the aggregate distributions were around $\alpha = 1.2$ for all the participating exchanges. Comparing these with values to ones from London ($\alpha = 1.5$) and Paris ($\alpha = 0.6$), it can be deduced that the trades in Nordic Exchanges believed more in significant price swings than Paris but less than London. This is because the higher the exponent, fewer orders are placed deep in to the book.

Table 5.1 *Power law decay exponents for all securities and exchange specific aggregates. In all cases* $x_{max} = 1000$.

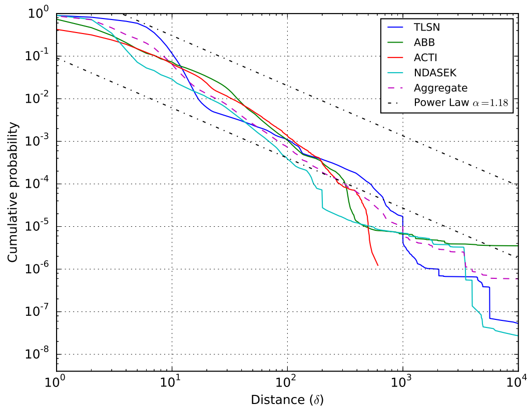
Helsinki		Stockholm		Copenhagen	
	α		α		α
NOK1V	1.146	TLSN	1.152	DANSKE	1.165
NES1V	1.155	ABB	1.171	MAERSKB	1.191
SAMAS	1.133	ACTI	1.252	VWS	1.178
FIA1S	1.296	NDASEK	1.173	BAVA	1.297
Aggregate	1.15	Aggregate	1.18	Aggregate	1.20

The power law is overtaken by more rapid decay (such as exponential) at some distance x_{max}. As can be seen from the Figures 5.1a, 5.1b and 5.1c, the exact distance from stock to stock. The more traded securities with thicker order books generally have higher x_{max}. The point where exponential decay over takes the power law is also affected by the tick size and the price $b(t)$, since the bid side order placement is limited by the price floor.

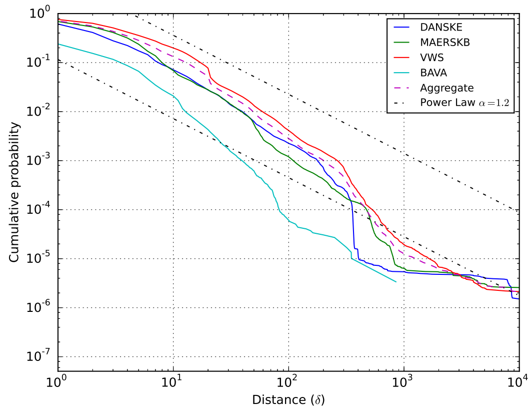
There are significant humps around relative prices corresponding to €1 distance, such as 500 ticks (with $\delta = 0.002$) for NOK1V and 100 (with $\delta = 0.01$) for SAMAS. These can be an indication of pegged orders with €1 offset or of algorithmic trading constantly posting flickering orders at these distances. Controlling for these outliers allows power law to provide a relative good fit for periods of about $\Delta \in [10, 800]$ even for less liquid stocks.



(a) Helsinki



(b) Stockholm



(c) Copenhagen

Figure 5.1 *Relative prices of new limit orders. Aggregated empirical distributions and their power law fits in dashed lines.*

Previous studies have attempted to generalize order placement in an exchange by providing just a one power law exponent. Exponents for fit with aggregate CDF is provided in Figure 5.1, but these aggregations are dominated by the high volume stocks such as NOK1V in Helsinki. In order to gain more explanatory power, a stock level or stock class (e.g. high, mid and low liquidity stock classes) exponents should be used. Additionally, the analysis is complicated by the relatively complicated tick setting procedure used by the Nordic Exchanges. This means that the connection with relative prices in ticks and monetary amounts is not static (what is more, OMX Nordic employed two different tick size tables during the time span of the sample).

Order sizes

Order size distributions have previously shown to be highly heterogeneous and to have heavy Pareto-type tails (Cont and Larrard 2012, pp. 14 – 15). The study found that order sizes themselves do not exhibit significant autocorrelation, but that the change in queues at different relative price levels have negative correlation with the opposite sides. Nevertheless, the order sizes are usually studied either as an independent distribution of random variable or as orders being of constant size.

Average order sizes are an important factor in the stochastic order book model by Cont et al. (2010), because all the order flow for a given security is modeled to consist of orders of a constant size (*unit size*). This assumption makes the model more simple but in empirical literature the order sizes have shown to vary greatly. Cont et al. use stock's average limit order size as the unit size. Nevertheless, the order size is an important consideration when order insertion is modeled and their distribution should be taken into account.

Table 5.2 *Incoming limit order size averages*

	$\bar{q}_l$	mode	median	σ
NOK1V	4663.41	1000	2094.48	9416.63
NES1V	654.32	400	400.55	1213.86
SAMAS	493.11	300	399	918.89
FIA1S	1183.64	1000	903.57	1699.71
TLSN	6922.03	2000	2800.35	9859.65
ABB	1807.47	1000	800.64	147179.3
ACTI	1052.61	1000	512.66	1714.33
NDASEK	5687.04	2000	2799.13	7478.6
DANSKE	741.48	400	461.83	1625.85
MAERSKB	2.37	1	1.63	26.85
VWS	721.81	100	399.23	1359.04
BAVA	497.31	100	340.01	650.44

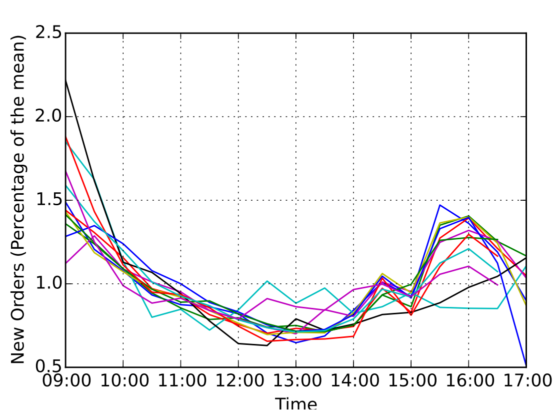
For example Zhao (2010), Alexander and Peterson (2007) and Harris and Panchapagesan (2005) have shown that traders prefer "even" order sizes, i.e. multiples of ten or five. This kind of behavior is challenging to capture with regular statistical distributions. Additional factors such as the total committed amount of money is also seen as an important consideration in new order insertion. For the purposes of CST-model, descriptive statistical values for analyzed securities are produced in Table 5.2.

5.1.2 Intra-day patterns

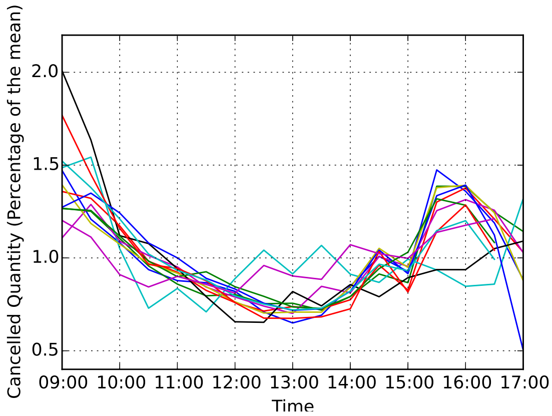
The trading and limit order insertion generally follows an *U-shaped pattern*, where the most of the activity is concentrated around market open and market close. In order to examine the seasonality in the data set, the trading day was split into half hour long periods. This is total of 17 per day for Helsinki and Stockholm and 18 for Copenhagen. Due to the last time span being only 25 minutes long, the relative values in the following results have been scaled accordingly.

Order insertions, cancellations and trades

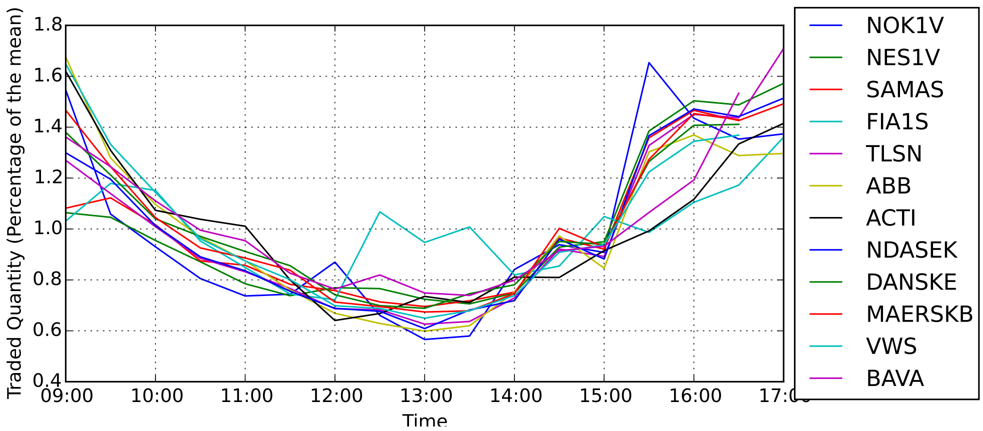
The general pattern of orders, cancellations and trades can be seen in Figures 5.2a, 5.2b and 5.2c. All the analyzed securities roughly have a common intra day pattern across exchanges, but it's more prominent in stocks with higher volume. For example FIA1S activity moves to a opposite direction compared to the aggregated values at 12:00 – 14:00.



(a) Orders



(b) Cancellations



(c) Trades

Figure 5.2 *Order, cancel and trade quantities for every half hour period*

Order and cancellation volumes follow the U-shape pattern apart from the last 30 minutes of continuous trading, where the values actually decrease. The activity seems to peak at 15:30 – 16:00 for every stock, but the effect is especially prominent for NOK1V. This is most likely due to NOK1V cross listing in New York Stock Exchange, market open of which is on 15:30 CET (9:30 EST). The NYSE and futures market opening seems to have smaller but noticeable effect on trading activity even if the security is not cross listed. (Halling et al. 2008) (Abhyankar et al. 1997)

From the three diurnal patterns studied, trading activity fits the U-shape the best, highest amount of trading happens around market open and close. It nevertheless has a hump is again around 15:30 – 16:00 correlating with other market opening times, but continues to increase until the market close.

This kind of correlation with foreign market opens have been noted previously in Switzerland by Ranaldo (2004). He observed that trading uncertainty increases during opening periods of U.S. markets, which in turn decreases the total number of outstanding orders (i.e. increases the cancellation rate) and limits the new order submissions near the best prices. Generally the traders want to *close* their positions before the increase in the amount of uncertainty and *take* new positions after it has waned. This hypothesis fits well with the observations from the Nordic Stock Exchanges.

Volatility and order book depth

Increased limit order flow thickens the order books, which reduce the possibility of short term price swing, while cancellations and market orders have an opposite function. Intra day price volatility patterns can be thus estimated via order book depths by using *price impact coefficients* in the Equation 3.13. This is done in the section 5.3 by leveraging the OFI model.

5.2 Event arrival rates

Limit order book event arrivals can be studied by using the distribution of inter-arrival times (or durations) and how they are affected by the preceding order flow. Order flow event durations can be exemplified by the use of constant arrival rates or more complex event arrival processes. In this section both the naïve statistical distribution fits and their parameters as well as the self-exciting property of order arrivals are examined.

5.2.1 Inter-arrival times

Inter-arrival times of the different types of order book events can be measured from the dataset as the difference of times corresponding events are handled by the matching engine. Since the ITCH feed represents the technical underpinnings of the limit order book, single event (e.g. order insertion) can lead to multiple events occurring with a same time stamp. Such events occur with the duration of *zero*.

Table 5.3 *Inter-arrival percentiles for order flow components ($\forall\Delta$). These show the order of magnitude differences between low and high liquidity stocks. Values in milliseconds.*

		NOK1V	NES1V	SAMAS	FIA1S	TLSN	ABB	ACTI	NDASEK	DANSKE	MAERSKB	VWS	BAVA
Orders	5 %	0	0	0	0	0	0	0	0	0	0	0	1.01
	10 %	0	0	0	1	0	0	2	0	0	0	0	12
	25 %	0	0	0	30	0	0.01	89	0	0.04	0	0	1361
	50 %	0.01	15	14	8347	0.06	10	2505	0.05	23.01	14	8	11575
	75 %	16	530	417	73776	31	174	18995	22.01	1001	395	250	56037
	99 %	2217	18281	10696	1185184	3684	4677	359110	2852	26646	18237	11571	629245
Cancels	5 %	0	0	0	0.03	0	0	0.04	0	0	0	0	5
	10 %	0	0	0	0.03	0	0	0.04	0	0	0	0	5
	25 %	0	0	0	27.02	0	0	28.01	0	0.02	0	0	1719
	50 %	0	10	7	5259	0	4	3130	0	21	10	3.01	13210
	75 %	1.01	484	314	65746	3	114	24800	0.1	987	306	210	63090
	99 %	1982	19057	11258	1686268	3091	4581	445054	2103	28969	19448	12335	720491
Trades	5 %	0	0	0	0	0	0	0	0	0	0	0	0
	10 %	0	0	0	0	0	0	0	0	0	0	0	0
	25 %	0	0	0	0	0	0	0	0	0	0	0	0
	50 %	0	22	21	25227	8	9	1762	0.01	28	33	26	13120
	75 %	737	6216	4826	338393	2362	5720	47222	1532	5865	6196	3069	175996
	99 %	37557	217794	142865	4056807	82432	141766	1173630	87808	177101	182916	102297	2527673

Limit and market orders

In order to examine the relations of orders that represent actual added quantity, not reinsertions due to partial trades or cancellations, only orders that are not updates were considered. Note that some types of dependent orders, such as pegged orders, can not reliably be filtered, but the number of such orders were assumed to be small. The trades and cancellations were considered as is, because they don't represent state information in the order book *per se*, but are instead immediately executed against active orders.

Inter-arrivals' cumulative distribution percentiles can be seen in Table 5.3. These are the inter-arrival rate of events of any size and relative price. The difference in liquidity is once again apparent and the clustering of different kind of events can also be detected. Considering the fact that the round trip times experienced by the market participants are in best case still in multiple milliseconds, the order insertion and cancellation decisions are generally not made in knowledge of the *actual* state of the order book.

The high probability of short inter-arrival times and the long tails would point them to fit to the Weibull distribution discussed in the Chapter 3. To test this, least squares fitting to was attempted for every stocks' whole data set's aggregated distribution. High volume stocks produced reasonable fits seen in Figure 5.3, while lower stocks with less order activity were more dominated by outliers.

For theoretical (e.g. stochastic) order book modeling conditionalized inter-arrival times are generally more interesting. For example in Cont et al. (2010) arrival times at certain distances of bids and asks are estimated to be exponentially distributed.

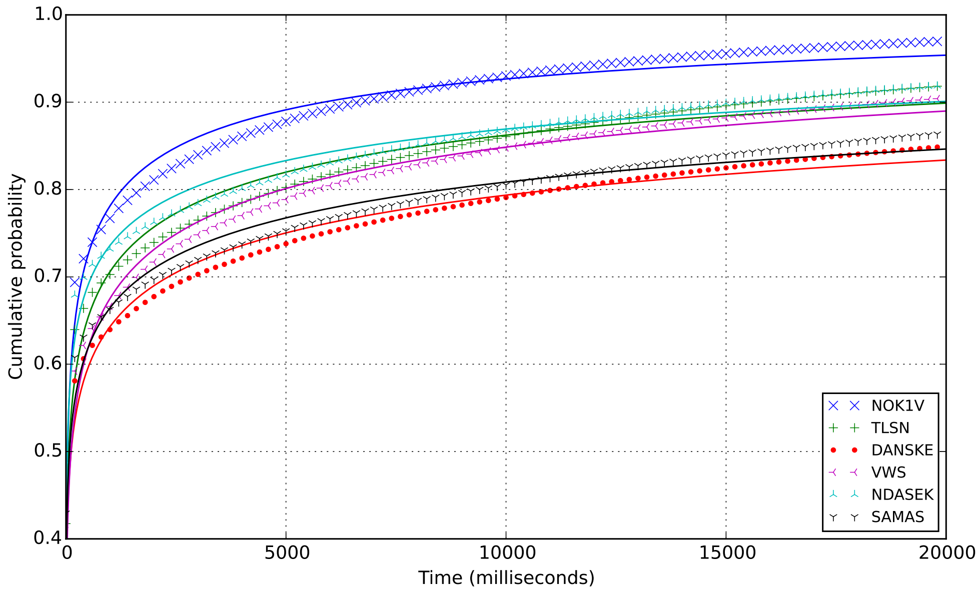


Figure 5.3 *Cumulative inter-arrival distributions using 200 ms bins ($\forall\Delta$). Solid lines are least squares fits of Weibull distributions.*

The inter-arrival times for every stock were extracted from both sides of the book for combined $\Delta_o \in [1,5]$. For the price levels $\Delta_o \geq 5$, CST-model assume arrival rates to be constant (i.e. $\lambda^L(5) = \lambda^L(6) = \lambda^L(7)...$).

According to the CST-model, market orders should also be exponentially distributed. ITCH feed data does not have explicit announcement of market orders but the trade data can be used for a similar effect. Afterwards exponential distributions were fitted using the least squares method for orders and trades. All exponents as per the Equation 3.3 can be seen in Table 5.4 and for NDASEK orders in and trades in Figure 5.5.

As can be seen in the table and the figures, the exponential distribution does not decay fast enough near $\Delta_o = 1$. The empirical distributions are considerably more concentrated very short time spans that can be accounted for by exponential decay. Additionally there are unexplained humps around 20 milliseconds, which can be hypothesised to correspond with DMA participant round-trip times caused by latency.

In the case of trades, the heavy concentration is partly due to use of trade data instead of market order data. A single market order can generate multiple trades with same time stamp if it's matched against multiple active limit orders. Likewise cancellations and re-insertions of pegged orders generate several order book events whenever the variable changes that the order were pegged onto. The amount of such

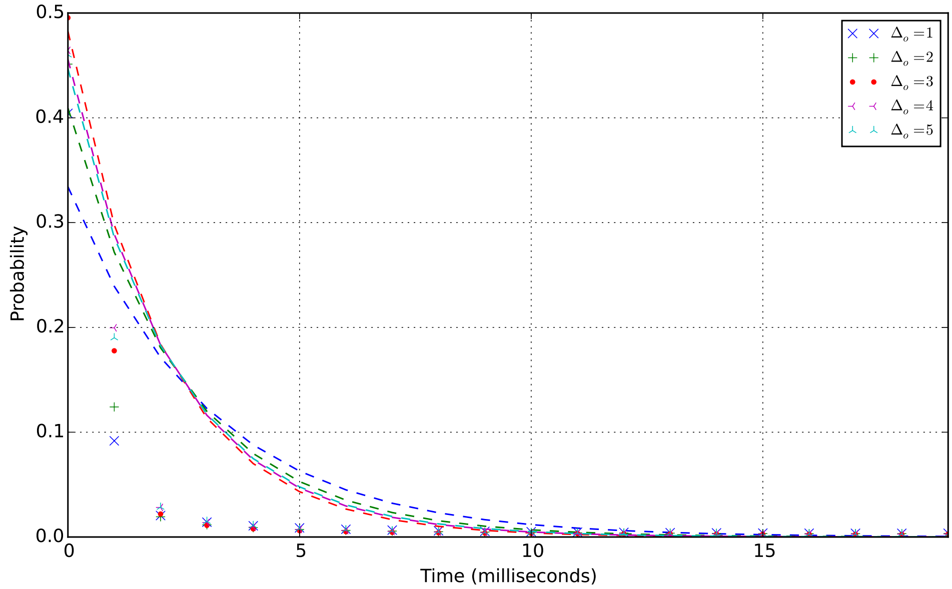


Figure 5.4 *Inter-arrival times for NDASEK orders. Exponential fits in dashed lines.*

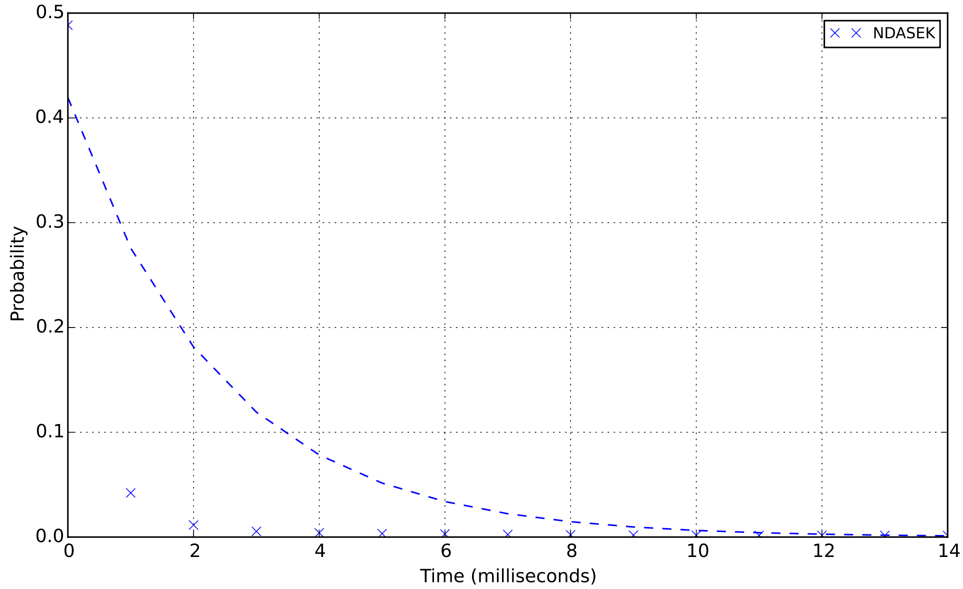


Figure 5.5 *Inter-arrival times for NDASEK trades. Exponential fit in dashed line.*

Table 5.4 *Inter-arrival exponential fit scale parameters for normalized distributions.*

	$\lambda^L(1)$	$\lambda^L(2)$	$\lambda^L(3)$	$\lambda^L(4)$	$\lambda^L(5)$	λ^M
NOK1V	0.3932	0.4504	0.5225	0.5048	0.4673	0.429
NES1V	0.1214	0.1723	0.1927	0.183	0.1966	0.333
SAMAS	0.1091	0.196	0.1891	0.1405	0.1401	0.3339
FIA1S	0.0001	0.0004	0.0071	0.0188	0.022	0.1403
TLSN	0.3185	0.3819	0.4094	0.3647	0.3621	0.3189
ABB	0.1681	0.1171	0.1481	0.1957	0.2579	0.3368
ACTI	0.0003	0.0006	0.0032	0.0087	0.0128	0.2294
NDASEK	0.3341	0.4092	0.4825	0.4557	0.4469	0.4193
DANSKE	0.0451	0.1205	0.1543	0.1429	0.1546	0.3265
MAERSKB	0.1475	0.2163	0.1862	0.1758	0.175	0.3162
VWS	0.0638	0.11	0.0966	0.1296	0.1739	0.3178
BAVA	0.0001	0.0001	0.0002	0.0003	0.0012	0.1407

events is still assumed to present only small part of the total order flow and the poor fits can not be reasonably explained by just semantical differences.

The results are very consistent with previous exponential studies conducted by Zhao (2010) and Toke (2011). Zhao found similar behavior studying limit order placement in *International Petroleum Exchange*. In the studies The exponential decay around small values ($\Delta_o \leq 5$) was not significant enough for both market orders and new order insertions. To improve agreement with empirical data, Toke studied multiple types of assets varying from stocks to Euribor futures. Both studies swapped Poisson process to Hawkes processes, where order flows consisted of periods of low and periods of high arrival rates, clustered in time.

Cancellations

As for cancellation rate $\theta(\Delta_o)$ Cont et al. proposes estimation method that uses average limit order size $\bar{q}_l$ and average cancellation size $\bar{q}_c$. Denoting total number of cancellations at distance Δ_o $N_c(\Delta_o)$ the model's cancellation rate can be estimated with

$$\theta(\Delta_o) = \frac{N_c(\Delta_o)}{TQ_{\Delta_o}} \frac{\bar{q}_c}{\bar{q}_l}, \tag{5.1}$$

for $\Delta_o \in]\infty, 5]$. T is the total sample length and Q_{Δ_o} average orders sitting at distance Δ_o. Due to the fact that in this paper we have access to individual cancellation data, a different procedure is proposed. Denoting total cancelled quantity a given distance C_{Δ_o} and noticing

$$C_{\Delta_o} \approx N_c(\Delta_o)\bar{q}_c, \tag{5.2}$$

we arrive at the equation

$$\theta(\Delta_o) = \frac{C_{\Delta_o}}{TQ_{\Delta_o}\bar{q}_l}, \tag{5.3}$$

which is the total cancellation rate at distance Δ_o. The time weighed average order book depth is used as Q_{Δ_o}. $\bar{q}_l$ can be acquired from Table 5.2. Note that this equation gives the cancellation rate assuming all the order be of average size (this is consistent with the model, Cont et al. defined rates for cancellations and orders of unit sizes). Cancellation rates for first 5 levels can be seen in Table 5.5

Table 5.5 *Single share cancellation rate constants for first five LOB levels.*

	$\theta(1)$	$\theta(2)$	$\theta(3)$	$\theta(4)$	$\theta(5)$
NOK1V	7.394	6.285	9.984	11.615	4.844
NES1V	2.149	1.433	1.042	0.960	0.843
SAMAS	3.752	1.792	1.023	0.978	0.889
FIA1S	0.105	0.052	0.020	0.015	0.011
TLSN	3.387	2.321	3.767	3.426	2.840
ABB	5.907	4.477	3.172	1.147	0.755
ACTI	0.294	0.164	0.093	0.086	0.068
NDASEK	4.436	5.437	10.751	8.826	5.470
DANSKE	1.328	0.848	0.498	0.329	0.277
MAERSKB	1.406	0.805	0.568	0.485	0.468
VWS	3.984	3.113	2.358	1.490	1.029
BAVA	0.101	0.032	0.017	0.014	0.012

Comparing these to the data set is difficult, since the rate is assumed to depend on the prevailing queue size. The empirical cancel rates have previously shown to have hump near the best price and a second further away. Relative price cancel rates additionally show a greater variance between stocks than orders. This is seen in the Figure 5.6, where cancel rates tick wise are produced for 6 stocks. (Potters and Bouchaud 2003)

The figure does not take into account the actual outstanding queues at different distances at the time of cancellation. The number of cancels are instead divided by the average number of shares at distance Δ. The rates presented in the figure are comparable to cancel rates previously studied by Potters and Bouchaud (2003, p. 136 – 137). As can be seen in the figure, the cancel rate is very low for orders at

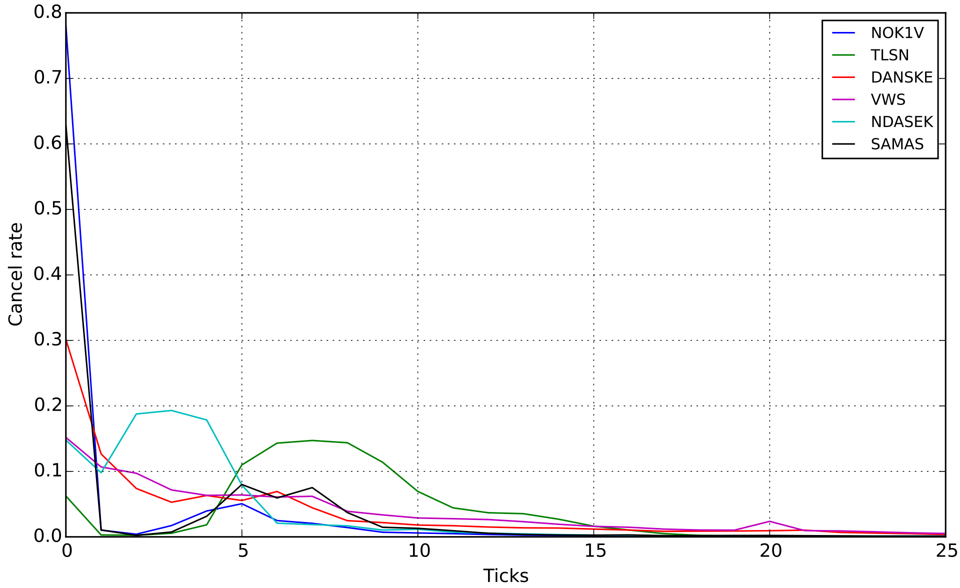


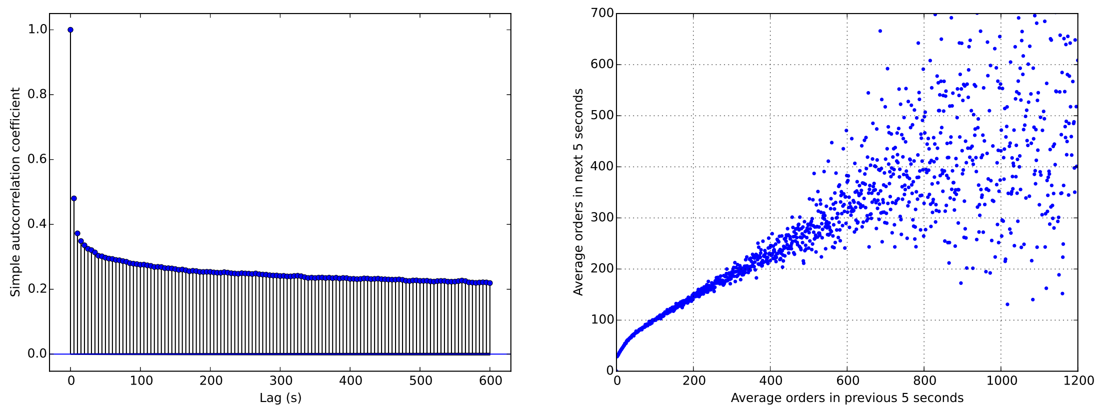
Figure 5.6 *Cancel amounts per relative price.*

relative price $\Delta = 1$, while the rate compared to the average depth is very high at the best prices. Another hump exists at relative price of 5 to 10 ticks. According to Potters and Bouchaud (2003) this is caused by the fact that most of the orders are cancelled either at the best price very quickly after insertion or after the price has started to move against it and the probability of execution has thus decreased.

5.2.2 Event clustering

While the previous section showed that the order book events and order flow inter-arrival times were heavily concentrated in short time spans and had long tails, the clustering of order book events can be further investigated. Recent limit order book models (Zhao 2010; Toke 2011) have assumed the order flow to have self-exciting property, i.e. increase in order flow attracts more order flow in the short term and vice versa. These effects require a more sophisticated than the regular homogeneous Poisson arrival processes.

In order to investigate clustering, the sample was split into 5 second periods and total number of different order book events during that period was calculated. Due to the size of the time series the autocorrelation estimation was done with a Fast Fourier Transform up to lag of 120 periods (10 minutes). The clustering of activity was also studied by comparing the nominal amount of events in 5 second time spans before and afterwards. Results for NOK1V shown in Figures 5.7 and 5.8. Similar results were acquired using time spans of 10 and 20 seconds.

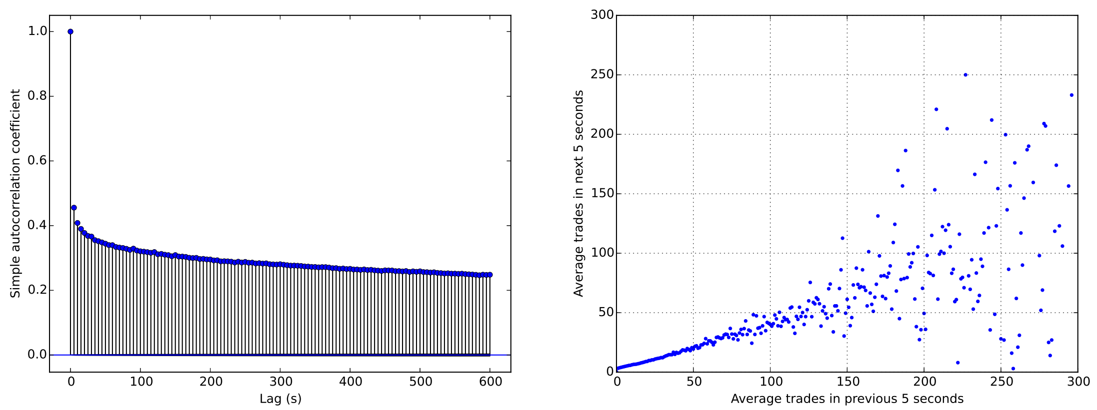


(a) Limit order autocorrelation

(b) Limit order feedback

Figure 5.7 *Autocorrelation and activity clustering for NOK1V limit orders (LL-effect)*

The number of new limit orders submissions and trades display significant autocorrelation at short lags, but reduces while still staying at 0.3 level 10 minutes afterwards. Similar results can be seen for the market order (trade) submissions. This fact holds for periods where there is less than 500 orders, for bigger orders the sample size is too small and the mean is dominated by outliers.



(a) Trade autocorrelation

(b) Trade feedback

Figure 5.8 *Autocorrelation and activity clustering for NOK1V trades (MM-effect)*

Similar tests can be conducted for other kinds of event clustering inter-excitability. For example market orders exciting limit order insertion has been previously studied and confirmed by Zhao (2010). These kinds of effects point to self-exciting properties of order flows that homogeneous Poisson processes of the CST-model can not account for.

These "market making" properties have been called LL and MM-effects by Zhao (2010). As can be seen from the Figures 5.7 and 5.8, the order flow is self exciting, when the total number of limit orders is not too high, after which the relaxation term takes over. For example for NOK1V LL-effect, the 5 second turning point is roughly 20 orders. Less than 20 limit orders have on average increasing number orders in the next period, where as periods with less than 20 orders have fewer orders in the following period. The effects were similar for all the stocks studied, but naturally the autocorrelation lags depended on the overall order activity levels. Turning points from self-exciting to self-relaxing behaviors were similarly liquidity dependent.

Self exciting and relaxing behavior intensities can be estimated by producing simple regressions to the figures such as 5.7b and 5.8b. In the cases, where the intensity feedback function is linear (such as MM-effect), only one regression is needed. The nonlinear feedback functions have both self exciting and relaxing properties, which requires individual regressions for both. For the self exciting regression, the scale coefficient $\beta_e > 1$ and for the relaxing $1 > \beta_r > 0$. In the special case of $\beta = 1$, the expected order activity for the following period is the same as for the current period. Another method of estimation would be the use of full kernel density estimates for the scale parameter of the intensity function.

5.3 Order Flow Imbalance

This section attempts to explore the assumptions and relations between Order Flow Imbalance, instantaneous price impact and volatility described by Cont et al. (2014). The calculations use 10 seconds as the shorter interval Δt and 30 minutes as the longer interval ΔT. The ITCH feed does not explicitly announce incoming market orders, but their effect on the top level of the book can be deduced from the trade data ($T = M$).

By equating trades with market orders, equations 3.8 and 3.10 can be combined to form the total Order Flow Imbalance for every Δt

$$\mathrm{OFI}_k = L_{b,k} - C_{b,k} - T_{b,k} - L_{s,k} + C_{s,k} + T_{s,k}, \tag{5.4}$$

where T_b and T_s are the traded quantities at the best bids and asks, respectively. The assumption that one market order generates just one trade is not strictly true, but due to the fact that the OFI is based on total quantities, not the number of distinct orders, the end result should be the same. Trades are just as good proxy for limit orders leaving the queue as market orders, when comparing the total quantity, not the total number of orders.

In order to assess the goodness of the OFI model and the corresponding coefficients, the midpoint price impact must be defined for a time span $t << T$. In this paper they are calculated from the data set by using the mid price difference in period Δt in ticks:

$$\Delta P_{k,i} = \frac{b(N(t_{k,i})) + a(N(t_{k,i})) - b(N(t_{k-1,i})) - a(N(t_{k-1,i}))}{2\delta}, \tag{5.5}$$

where $N(t_{k,i})$ is the last order book event in the period $[t_{k-1,i}, t_{k,i}]$. This is the change in $m(t)$ as a multiple of δ in a period Δt.

As discussed before the price impact function can be assumed to be linear. As per the Equation 3.13, $\hat{\beta}_i$ can be estimated for the period ΔT from the empirical data by using the linear regression

$$\Delta P_{k,i} = \hat{\alpha}_i + \hat{\beta}_i \,\mathrm{OFI}_{k,i} + \hat{\epsilon}_{k,i}. \tag{5.6}$$

Apart from using the full order, trade and cancellation data to derive OFI not just the level 1 order book state, the procedure described follows the outline set by Cont et al. (2014, pp. 8 – 10), which has a more detailed discussion of the reasoning and associated caveats.

The data set was split to Δt long time spans for which the OFI was calculated. At the best case this would have yielded $783 \times 51 = 39933$ observations for Helsinki and Stockholm and $783*51 = 37584$ for Copenhagen, but time periods for which OFI was undefined (i.e. no LOB events) were left out of consideration. This is most apparent for low liquidity stocks, for which the order book events are highly clustered in time.

The $\hat{\beta}$ and $\hat{\alpha}$ were assumed to be constant for the period ΔT. Using a ordinary least squares regression to estimate the coefficients of Equation 3.13. Intercept, slope and determination coefficients averaged stock wise across the whole sample are available in Table 5.6

The model performs relatively well for high volume stocks such as NOK1V, TLSN and VWS, but the fit is generally poorer than reported by Cont et al. (2014). Their mean R^2 across stocks sampled from NASDAQ New York was 65 % while for this study's data set it's 47 %. The R^2 values for lower liquidity stocks are significantly worse, e.g. the model explains only 35.7 % of the observed variance in BAVA. Similar case of better performance with high volume stocks was reported by Cont et al., but the difference was not as significant. These differences can be due to larger number

Table 5.6 *Average values for the OFI model by security*

	α	β	R^2
NOK1V	0.000723	0.000015	0.572723
NES1V	0.000829	0.000175	0.520058
SAMAS	0.000387	0.000156	0.544211
FIA1S	0.065854	0.000728	0.488405
TLSN	-0.000674	0.000043	0.546146
ABB	0.001565	0.000022	0.429070
ACTI	-0.025881	0.000338	0.376004
NDASEK	0.000266	0.000014	0.514604
DANSKE	-0.000486	0.000135	0.402650
MAERSKB	-0.001268	0.043084	0.415677
VWS	0.005646	0.000294	0.461659
BAVA	-0.001789	0.000742	0.357090

stocks analyzed by Cont et al. (2014) or general lower liquidity in the Nordic Stock Exchanges, which leads to a more unstable $\hat{\beta}$. This would make the assumption that $\hat{\beta}$ stays constant for half hour questionable in case of thinner order books. The intercept values were not found to be significant.

The values of $\hat{\beta}$ has been previously shown to have a significant seasonality by Cont et al. (2014). This is implied in the regression by the assumption that the coefficients $\hat{\alpha}$ and $\hat{\alpha}$ stay constant for the longer ΔT (in this case, 30 minute) duration. To explore this, the values of $\hat{\beta}$ for every half hour were averaged across all the stocks (see Table 5.7). The values are estimates of mean temporal midpoint price variance caused by the difference of order flow at the best prices. The coefficients are not directly comparable between different classes of stocks, due to the differing tick size derivation tables.

Table 5.7 *Averaged values for $\hat{\beta}_k$ every continuous trading half hour in HEL, STO and CPH. Last row of values are undefined for CPH due to shorter continuous trading period.*

	NOK1V	NES1V	SAMAS	FIA1S	TLSN	ABB	ACTI	NDASEK	DANSKE	MAERSKB	VWS	BAVA
09:00	0.000022	0.000259	0.000258	0.000407	0.000071	0.000033	0.001605	0.000022	0.000209	0.056215	0.000402	0.001288
09:30	0.000018	0.000230	0.000196	0.002358	0.000057	0.000026	0.000314	0.000017	0.000160	0.048931	0.000337	0.001567
10:00	0.000017	0.000201	0.000180	0.000226	0.000048	0.000024	0.000208	0.000015	0.000148	0.044734	0.000314	0.000943
10:30	0.000016	0.000190	0.000168	0.000342	0.000046	0.000022	0.000329	0.000014	0.000138	0.044931	0.000298	0.001123
11:00	0.000015	0.000181	0.000157	0.000774	0.000044	0.000022	0.000389	0.000014	0.000138	0.042819	0.000291	0.002412
11:30	0.000015	0.000173	0.000151	0.000254	0.000042	0.000022	0.000311	0.000013	0.000133	0.041933	0.000287	0.000577
12:00	0.000015	0.000171	0.000148	0.003770	0.000042	0.000022	0.000190	0.000014	0.000124	0.042984	0.000283	0.001200
12:30	0.000015	0.000169	0.000147	0.001424	0.000040	0.000022	-0.000315	0.000013	0.000124	0.040629	0.000272	0.000710
13:00	0.000015	0.000172	0.000144	0.000461	0.000041	0.000021	0.000300	0.000013	0.000124	0.040760	0.000266	0.000492
13:30	0.000015	0.000172	0.000142	0.000218	0.000040	0.000021	0.001593	0.000013	0.000125	0.039840	0.000266	0.001377
14:00	0.000015	0.000179	0.000154	0.001039	0.000040	0.000022	-0.001116	0.000013	0.000126	0.041000	0.000276	0.001472
14:30	0.000015	0.000172	0.000149	0.000431	0.000040	0.000021	0.000168	0.000013	0.000130	0.041383	0.000277	-0.001546
15:00	0.000015	0.000156	0.000137	0.000170	0.000037	0.000020	0.000053	0.000012	0.000123	0.038068	0.000267	0.001097
15:30	0.000016	0.000153	0.000141	0.000307	0.000041	0.000022	0.000250	0.000013	0.000128	0.041804	0.000284	0.000819
16:00	0.000014	0.000141	0.000133	0.000231	0.000039	0.000021	0.000252	0.000012	0.000123	0.038937	0.000283	0.000393
16:30	0.000013	0.000131	0.000123	0.000369	0.000036	0.000019	0.000313	0.000012	0.000101	0.043750	0.000295	-0.001185
17:00	0.000012	0.000121	0.000118	0.000133	0.000035	0.000020	0.000227	0.000012	–	–	–	–

The high variance of $\hat{\beta}_k$ in low volatility stocks is an another indication that the OFI model can not fully explain midpoint price movements of stocks with under thousand order book events per day on average. This difference between high and low liquidity stocks statistical behavior has been explored by Cont and Larrard (2012), who defined the heavy traffic limits for limit order book markets.

On the other hand, high liquidity stocks' $\hat{\beta}_k$ mostly follow the diurnal behaviour outlined previously by Cont et al. (2014, pp. 15 – 16) as can be seen in Figure 5.9). Higher values for $\hat{\beta}_k$ indicate that incoming orders can more easily affect $m(t)$. This correlates with previous studies where the market depth was found to be shallower in the morning and thus more easily movable by new active limit order insertions or deletions (Lee et al. 1993). The high number of trades and low number of new limit orders in last three periods is compensated by the drop in cancelled quantities.

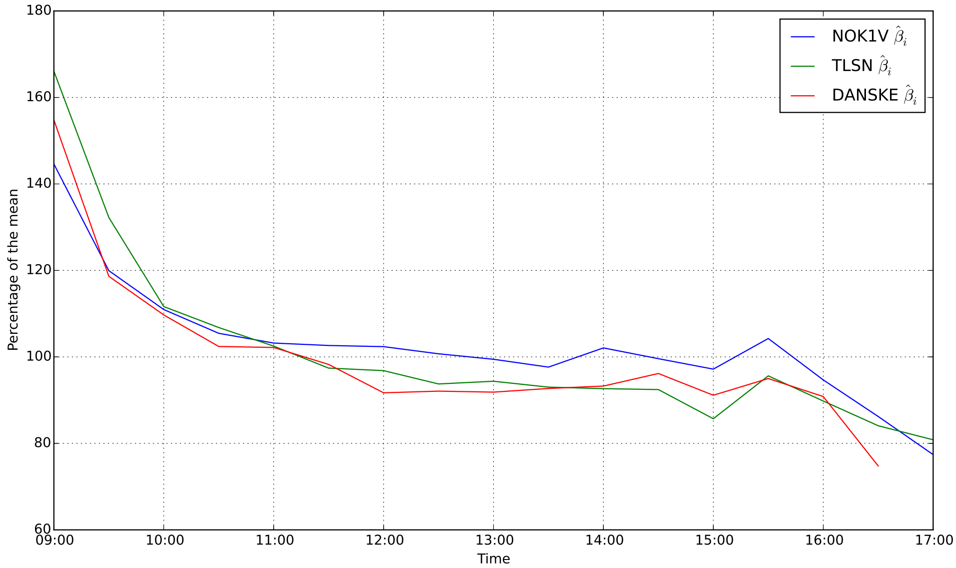


Figure 5.9 *Average $\hat{\beta}$ for liquid securities NOK1V (HEL), TLSN (STO) and DANSKE (CPH)*

Comparing the diurnal behavior of price impact (or similarly, price volatility) with the diurnal periodicity of limit order book events discussed previously in the Chapter 5.1, the $\hat{\beta}$ roughly follows the general activity levels across the day. The high number of cancellations around 14:00 – 15:00, contribute to thinning of the order book and increase the possible midpoint price movement caused by the Order Flow Imbalance. This intra-day increase in $\hat{\beta}$ was not observed in Cont et al. (2014), because the study used data from the New York Stock Exchange. The reason for the increase seems to be the cancellation of active limit orders at times just before New York Stock Exchange opens. As mentioned before, this is can be seen as an indication of increased trading uncertainty. The effect apparent even for stocks that are not cross listed in multiple exchanges.

6. CONCLUSIONS

The main research problems of this thesis can be split into two parts: "How order flow dynamics can be modeled and how these models are calibrated?"and "What are these calibration values for Nordic Stock Exchanges?". The critique and assessment of the assumptions made by the theoretical models has overlaps with both of these questions. Additionally, the calibration values themselves have economic implications that can be used to comprehensively characterize the NASDAQ OMX Nordic limit order books.

These statistical event level variables of Nordic Stock Exchanges exhibit significant similarity with the ”stylized facts” reported in previous studies. The length of the data set used allows the study to reliably specify and describe these characteristics for NASDAQ OMX Nordic. Additionally, the event-by-event time series allowed to explore the facts without making additional assumptions that could have affected results. As such, the thesis is a very comprehensive assessment of critical For example, the power law tails of relative order placement, the study reported decay exponents of $\alpha = 1.15$ in Helsinki, $\alpha = 1.18$ in Stockholm and $\alpha = 1.20$ in Copenhagen. This places the studied exchanges between LSE and Paris Bourse in trader expected volatility.

The significant seasonality in order book event activity was observed in all the exchanges studied. There was some differences between stocks based on the liquidity, but generally the activity followed the previously described u-shaped pattern. Moreover, differences arose during the North American market opens around 14:00 – 15:00 CET, where the order flow significantly was affected by this outside effect. These were also the moments, when the intraday order book were at their thinnest. This is not a behavior completely unique to Nordic Stock Exchanges, as it has been observed in other smaller exchanges, such as Switzerland. Most of significant empirical LOB studies have been done in exchanges, where this effect is less prominent, such as LSE or NYSE.

Finally, defining the characteristic values of the exchange’s order flow allows exploring the nature of limit order market and enables one to make inferences from such facts as expected volatility, but the most future applications depend on combining the empirical facts with the theoretical models. Both empirical literature and sophisticated theoretical LOB modeling has gained increased attention during the past few years. Still there is a fundamental disconnect between the empirical observations and well defined, testable mathematical models. Recent models might be able to produce quantifiable and applicable results for shorter time spans, but fail to generalize themselves and fit long term limit order book dynamics. Some of the simplified models such as Order Flow Imbalance can explain some variables well and have number of applications. In this paper OFI model attained R^2 of over 0.4 for high liquidity stocks. Nevertheless, such simplified models still fail to bridge the market microstructure and long term variables.

While these Zero Intelligence models perform relatively well at shorter time scales, they are still in some respect simplified ”toy models”. The assumptions of constant order sizes and arrival rates are not supported by this thesis or the previous academic literature. Long term order flow is clearly more affected by strategic decisions of participants and can't be captured reliably by simple statistical distributions. This means that the future developments in limit order book dynamics should incorporate the event clustering and activity self excitement / relaxation. This can be done be either combining the Zero Intelligence with Dynamic Equilibrium models or introducing variable arrival rates to the Zero Intelligence modeling. This study reported the arrivals to be highly clustered in time and exhibiting significant autocorrelation at short time lags (The LL and MM effects). The distribution of arrival times decreased monotonically, with humps around 20 milliseconds.

The variable arrival rates have been studied by Zhao (2010), where as the combining of ZI and DE models has been attempted in Agent Based Models (ABM), where heterogeneous agents interact according to specified individual rules. Chakraborti et al. (2011) explored the models have found the promising compromise between the economic ”realism” of DE and process calibration of ZI models. These models might be able to capture and explain the longer term dynamics, while simplified ZI models have more applications due to their easy computability and ”good enough” estimations for the shorter intraday dynamics.

On the empirical or statistical side, the most important research questions are incorporating the very large data sets nowadays available into the analysis. The inferring of detailed market variables from low level data, such as TAQ, is no longer needed since the availability of event-by-event LOB time series. The amount of ”noise” is also expected to grow because of the increase in high-frequency and algorithmic trading, which places constraints on inferring LOB's characterising values. Developments in such analysis would also have significant regulatory implications and would enable more deep study of LOB market ”health”. All in all, the study of limit order markets is in an interesting position, where the rapidly improving models and vast data sets hopefully allows us to answer these outstanding questions in the not too distant future.

BIBLIOGRAPHY

Abhyankar, A. et al. (1997). ”Bid-ask Spreads, Trading Volume and Volatility: Intraday Evidence from the London Stock Exchange”. *Journal of Business Finance & Accounting* 24.3, pp. 343–362.

Al-Suhaibani, M. and L. Kryzanowski (2000). ”An exploratory analysis of the order book, and order flow and execution on the Saudi stock market”. *Journal of Banking & Finance* 24.8, pp. 1323–1357.

Alexander, G. J. and M. A. Peterson (2007). ”An analysis of trade-size clustering and its relation to stealth trading”. *Journal of Financial Economics* 84.2, pp. 435–471.

Alstott, J., E. Bullmore, and D. Plenz (2014). ”powerlaw: a Python package for analysis of heavy-tailed distributions”. *PLoS One* 9.1, e85777.

Bak, P., M. Paczuski, and M. Shubik (1997). ”Price variations in a stock market with many agents”. *Physica A: Statistical Mechanics and its Applications* 246.3, pp. 430–453.

Bisiere, C. and T. Kamionka (2000). ”Timing of orders, orders aggressiveness and the order book at the Paris bourse”. *Annales d’Economie et de Statistique*, pp. 43–72.

Bouchaud, J.-P., M. Mezard, and M. Potters (2002). ”Statistical properties of stock order books: empirical results and models”. *Quantitative finance* 2.4, pp. 251–256.

Chakraborti, A. et al. (2011). ”Econophysics: agent-based models”. *Quantitative Finance*, Quantitative Finance, Vol. 11, No. 7, July 2011, 1013–1041 Econophysics.

Challet, D. and R. Stinchcombe (2001). ”Analyzing and modeling 1+1d markets”. *Physica A: Statistical Mechanics and its Applications* 300.1, pp. 285–299.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). ”Power-law distributions in empirical data”. *SIAM Review* 51.4, pp. 661–703.

Cont, R. (2011). ”Statistical modeling of high-frequency financial data”. *Signal Processing Magazine, IEEE* 28.5, pp. 16–25.

Cont, R. and A. D. Larrard (2012). ”Order book dynamics in liquid markets: limit theorems and diffusion approximations”.

Cont, R., S. Stoikov, and R. Talreja (2010). ”A stochastic model for order book dynamics”. *Operations research* 58.3, pp. 549–563.

Cont, R., A. Kukanov, and S. Stoikov (2014). ”The price impact of order book events”. *Journal of financial econometrics* 12.1, pp. 47–88.

Farmer, J. D., P. Patelli, and I. I. Zovko (2005). ”The predictive power of zero intelligence in financial markets”. eng. *Proceedings of the National Academy of Sciences of the United States of America* 102.6, pp. 2254–2259.

Foucault, T. (1999). "Order flow composition and trading costs in a dynamic limit order market". *Journal of Financial markets* 2.2, pp. 99–134.

Foucault, T., O. Kadan, and E. Kandel (2005). "Limit Order Book as a Market for Liquidity". English. *The Review of Financial Studies* 18.4, pp. 1171–1217.

Glosten, L. R. (1994). "Is the electronic open limit order book inevitable?" *The Journal of Finance* 49.4, pp. 1127–1161.

Gould, M. D. et al. (2013). "Limit order books". *Quantitative Finance* 13.11, pp. 1709–1742.

Gu, G.-F., W. Chen, and W.-X. Zhou (2008). "Empirical regularities of order placement in the Chinese stock market". *Physica A: Statistical Mechanics and its Applications* 387.13, pp. 3173–3182.

Halling, M. et al. (2008). "Where is the market? Evidence from cross-listings in the United States". *Review of Financial Studies* 21.2, pp. 725–761.

Harris, L. E. and V. Panchapagesan (2005). "The information content of the limit order book: evidence from NYSE specialist trading decisions". *Journal of Financial Markets* 8.1, pp. 25–67.

Hautsch, N. and R. Huang (2011). "Limit order flow, market impact and optimal order sizes: evidence from NASDAQ TotalView-ITCH data".

Hollifield, B., R. Miller, and P. Sandas (2004). "Empirical analysis of limit order markets". *The Review of Economic Studies* 71.4, pp. 1027–1063.

Jain, P. K. (2003). "Institutional design and liquidity at stock exchanges around the world".

Kaldor, N. (1957). "A model of economic growth". *The economic journal*, pp. 591–624.

Kukanov, A. (2013). *Stochastic Models of Limit Order Markets.*

Kyle, A. S. (1985). "Continuous auctions and insider trading". *Econometrica: Journal of the Econometric Society*, pp. 1315–1335.

Lee, C. M., B. Mucklow, and M. J. Ready (1993). "Spreads, depths, and the impact of earnings information: An intraday analysis". *Review of Financial Studies* 6.2, pp. 345–374.

Lo, A. W., A. C. MacKinlay, and J. Zhang (2002). "Econometric models of limit-order executions". *Journal of Financial Economics* 65.1, pp. 31–71.

Lorenz, J. and J. Osterrieder (2009). "Simulation of a limit order driven market". *The Journal Of Trading* 4.1, pp. 23–30.

Maslov, S. (2000). "Simple model of a limit order-driven market". *Physica A: Statistical Mechanics and its Applications* 278.3, pp. 571–578.

Menkveld, A. and M. Zoican (2014). *Need for Speed? Exchange Latency and Market Quality.*

NASDAQ (2010). *NASDAQ OMX Nordic Market Model 1.1.*

NASDAQ (2011). *NASDAQ OMX Nordic Market Model 2.2.*

Parlour, C. A. (1998). "Price dynamics in limit order markets". *Review of Financial Studies* 11.4, pp. 789–816.

Parlour, C. A. and D. J. Seppi (2008). "Limit order markets: A survey". *Handbook of financial intermediation and banking* 5, pp. 63–95.

Potters, M. and J.-P. Bouchaud (2003). "More statistical properties of order books and price impact". *Physica A: Statistical Mechanics and its Applications* 324.1, pp. 133–140.

Ranaldo, A. (2004). "Order aggressiveness in limit order book markets". *Journal of Financial Markets* 7.1, pp. 53–74.

Rinne, H. (2008). *The Weibull distribution: a handbook.* CRC Press.

Rosu, I. (2009). "A dynamic model of the limit order book". *Review of Financial Studies.*

Rosu, I. (2014). "Liquidity and information in order driven markets".

Smid, M. (2015). "Estimation of Zero Intelligence Models by L1 Data".

Toke, I. M. (2011). "“Market making“ in an order book model and its impact on the spread". Econophysics of Order-driven Markets. Springer, pp. 49–64.

Zhao, L. (2010). "A model of limit-order book dynamics and a consistent estimation procedure".

Zovko, I. and J. D. Farmer (2002). "The power of patience: a behavioural regularity in limit-order placement". *Quantitative finance* 2.5, pp. 387–392.